

## **SESSION**

# **NOVEL METHODS AND SYSTEMS TO ENHANCE DIAGNOSIS + TOOLS FOR HOSPITAL EMERGENCY ROOMS + GOOD PRACTICES**

**Chair(s)**

**TBA**



# Modeling of Clinical Practice Guidelines for Interactive Assistance in Diagnostic Processes

P. Philipp<sup>1</sup>, Y. Fischer<sup>1</sup>, D. Hempel<sup>2</sup>, and J. Beyerer<sup>1</sup>

<sup>1</sup>Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Karlsruhe, Germany

<sup>2</sup>Steinbeis-Transfer-Institut Klinische Hämatookologie, Donauwörth, Germany

**Abstract**—Clinical Practice Guidelines (CPGs) include recommendations for actions and therefore provide a frame of reference for the medical practitioner during diagnostic processes. To facilitate the implementation of this recommendations we propose an interactive assistance. For this reason the CPGs of Chronic Myeloid Leukemia (CML) and Myelodysplastic Syndromes (MDS) are modeled using UML activities and Bayesian networks. To lower the barriers of CPG implementation we establish an interface between experts and models. It is based on translation rules transferring UML activities to Bayesian networks. The resulting models are used to provide an innovative assistance function allowing intuitive deviations from given CPG recommendations.

**Keywords:** Clinical Practice Guidelines, Diagnosis, Interactive Assistance

## 1. Introduction

A challenge of the medical diagnostic today can be seen in the reviewing and evaluation of the huge amount of publications. For example: A search of the term “myelodysplastic syndrome diagnosis” in Pubmed [1] yields about 12000 results. For a decision maker a profound inquiry concerning a specific topic can therefore be extremely laborious. Moreover, the steady growth of published findings makes it almost impossible for an individual to keep their knowledge up-to-date [2].

With the use of Clinical Practice Guidelines (CPGs) the consolidated medical knowledge can be condensed into general recommendations of actions. High quality CPGs are proposed by bodies of experts with respect to the current state of research [3]. Consequently, for an individual medical practitioner, CPGs open up a scope of actions and decisions in the context of a contemporary diagnostic practice.

In addition to the development and dissemination of a CPG (see Figure 1) the actual implementation of recommendations by the medical partitioner plays a decisive role. The physician has to adapt a diagnostic algorithm to the given boundary conditions (patient, equipment, medical experience) [4], [5]. Consequently there is a gap between theoretical knowledge and practical solutions. Additionally, barriers can arise from (for example) low acceptance of CPGs on the part of the physician. The described situation is subject of many researches [6], [7], [8], [9].

A passive dissemination (e.g. distribution via print media) has only little effect on the actual practitioners behavior [4]

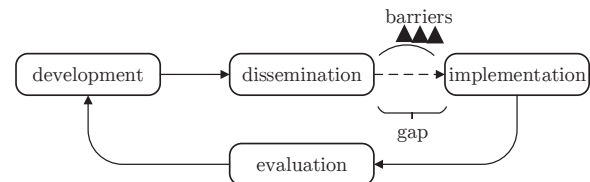


Fig. 1: Stages of a Clinical Practice Guide (CPG). The recommendations are development and disseminated. There is a gap between theoretical knowledge and practical actions performed by the medical practitioner. Additionally there can be barriers that have to be overcome.

[5]. Therefore we propose an interactive assistance of the practitioner during the diagnostic process which helps to reduce the gap between theoretical knowledge and practical solutions and helps to overcome barriers.

Barriers can arise from different CPG stages [10]. We are convinced that most of them could be moderated by creating the possibility to modify the recommendations by the practitioner himself. This includes for example the reduction of fear of regimentation or short-term modifications due to medical symposia etc. That’s why our approach of modeling medical knowledge comprises a dialogue between technical and medical domain experts as well as the modification of knowledge by the medical expert himself.

## 2. Approach of Modeling CPGs

Usually CPGs contain knowledge in form of texts and schematic diagrams that can not automatically be translated into models without further ado. We believe that the formalization of knowledge can be done via an expert dialog (see Figure 2). Therefor experts from the medical- and from the technical domain are necessary. Together they develop a CPG model in the form of a UML activity. Alternatively, the guideline can be interpreted by the medical expert, only. This approach is useful if a practitioner wants to modify an already existing UML activity independently. This bypass could help to lower barriers during the CPG implementation.

The benefit of our approach is that it is based on a UML activity. Different models representing the CPG can automatically be generated by only one given activity. Therefore, we developed translation rules. The UML activity in Figure 2 serves as an interface for the actual models used for providing assistance functions. These functions propose suitable examination values to the practitioner during the diagnostic process.

The automatic translation of a UML activity into a Petri net is based on the work of Störrle et. al [11], [12]. A

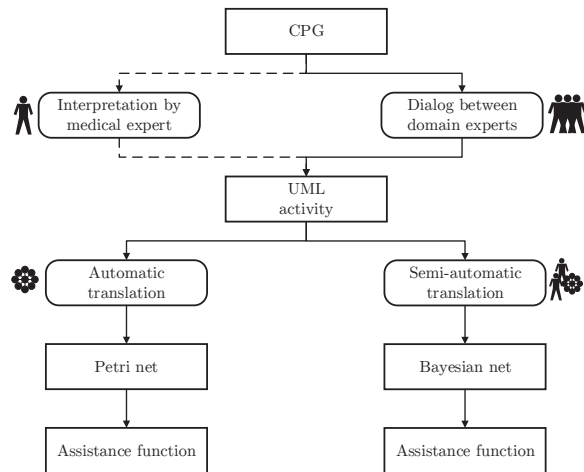


Fig. 2: A CPG is transformed into a UML activity by an expert's dialog. A bypass of the procedure is introduced to allow an exclusive modification by the medical expert. The UML activity serves as an interface to the models used for providing the actual assistance functions. These models are (semi-) automatically generated from the UML activity by translation rules.

new approach is used to semi-automatically translate a UML activity into a Bayesian net. What makes this translation semi-automatic is the fact that the parameters of the net (i.e. conditional probabilities) have to be assessed manually once the net structure is automatically generated from the activity. In this paper we focus on the translation of UML activities into Bayesian nets.

### 3. Construction of the Interface

UML activities have been chosen as an interface because their syntax is formalized and analyzed by various experts [13]. Furthermore, UML is accepted in software industry worldwide [13], [14], [15]. A huge benefit of UML activities is their easy comprehensibility for the medical- as well as the technical domain experts. This is a necessary precondition to make the experts' dialog work smoothly and to allow modifications by the medical expert on his own.

#### 3.1 State of the Art

The UML 2.4 comprises a total of 14 different chart types which can be divided into structural and behavioral diagrams [14]. Activity diagrams are among the latter and thus model not the static but the dynamic behavior of a system and its components. Thus, an activity answers the question of how a particular process or algorithm proceeds [14]. Control flows, object flows, actions, decisions and forks can be used to specify such an activity [15].

Figure 3 shows the typical routings which appear in the guideline models Chronic Myeloid Leukemia (CML) and Myelodysplastic Syndromes (MDS). The black dot represents the start of an activity (initial node), whereas the double circle corresponds to the end of an activity (activity final). The rounded rectangles are the actions that are to be performed, while the arrows represent the flow. Subfigure a) shows a case where actions are carried out one after another (sequential). Subfigure b) depicts a selective routing. Only

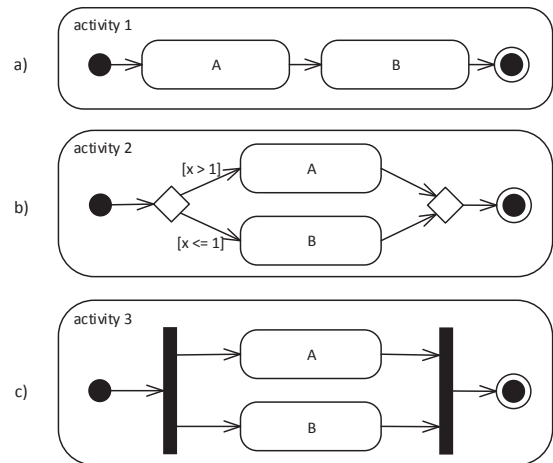


Fig. 3: The subfigures show the 3 typical routings appearing in the CPG models of CML and MDS. In a) the actions A and B are sequentially performed (one after another). In b) there is a decision to be made in order to perform either A or B. c) shows a routing where A and B are performed concurrently. Consequently both actions are performed in any possible order.

one of the two actions A and B is performed. The decision is represented by a diamond (decision node) and depends in this example on a variable  $x$  which is either greater 1 or not. The second diamond is called merge node as it merges the two possible flows. Subfigure c) shows a case where two actions can be performed concurrently. The flow is split up by a so called fork node (black bar). The join node on the right synchronizes the flows. As a consequence the flow continues, i.e. the activity ends, only iff both actions have been performed (in any arbitrary order).

#### 3.2 Modeling of a Particular Disease

Figure 4 shows the UML activity for CML. Due to the complexity of the CML CPG and the resulting size of the corresponding UML activity, the sketch emphasizes some parts of interest (magnifying glasses).

Subfigure a) depicts a sequence of actions. The first action is "suspicion of CML". That's an important precondition, since there are different CPGs for different diseases. Consequently the diagnosis algorithms of a CPG normally start with a specific suspicion for the disease under consideration. The second action is anamnesis (case history). During the anamnesis the practitioner asks the patient if he feels bone pain, i.e., the value "bone pain" is a result of the activity "anamnesis". We used a so called pin notation to specify a value as an output parameter for a given action [16]. The third action "physical examination" generates the output "spleen size". Actions can have several output- or input parameters.

Subfigure b) shows the action "Verify CML". This action involves some kind of assessment. Namely, the practitioner has to decide whether or not the examination values are proofing the disease. Therefore the decision if a disease is present is not modeled in a deterministic way (i.e. not by fixed rules). The final decision is up to the medical expert.

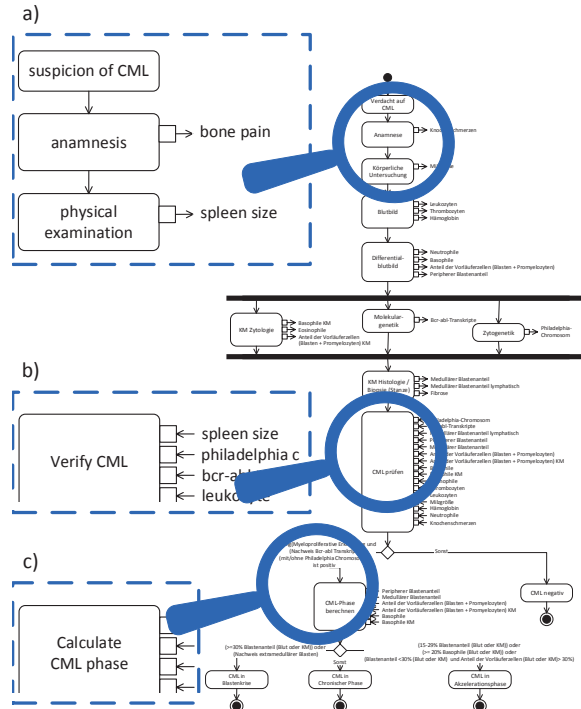


Fig. 4: In the background of this figure the UML activity for CML is shown. Because of the size of the activity the sketch emphasizes some parts of interest (magnifying glasses). Subfigure a) shows a sequential order of actions. Subfigures b) and c) show two different types of actions. The first one is a decision based on examination values that cannot be made by deterministic rules. Consequently the medical expert has to decide what to do - not the model. The second action is a decision which is based on a fixed rule (e.g. thresholds for particular blood test results).

Another decision is shown in Subfigure c). The keyword “calculate” emphasizes that this decision can be made on the basis of a specific rule. In this example, which type of CML is present can be derived by evaluating fixed rules (e.g. thresholds for particular blood test results).

The activity for MDS is about twice the size of the CML activity. The MDS activity is not shown in this paper, since the basic underlying concepts are the same as for CML. However, one important difference between the two diseases CML and MDS exists: In case of CML the practitioner is searching for examination values (e.g. blood test values) that proof the presence of the disease. MDS in contrast is a diagnosis of exclusion, i.e. rather than proofing MDS, known differential diagnoses of MDS are excluded. If all of the differential diagnoses are excluded, it is assumed to be proven that MDS is present<sup>1</sup>.

### 4. Bayesian Nets

Medicine is a famous area of application for Bayesian networks [17]. They are well suited for diagnostic processes because the probability for a diagnosis can be calculated by

<sup>1</sup>This modus operandi is known from quizshows like “Who Wants to Be a Millionaire?” – i.e. by excluding 3 of possible 4 answers the candidate is able to deduce the right answer.

successively adding examination values. Using this approach the practitioner can complete fragmentary examination values independent of a fixed sequence of tests.

### 4.1 State of the Art

Bayesian networks are probabilistic graphical models (PGMs) since they combine graph theoretic approaches with approaches of probability theory. They can be used to represent a probability distribution more compactly by taking advantage of independencies between variables. For the representation of these independencies the representation as a graph seems to be natural.

A Bayesian network consists of a probability distribution P and a graph  $G = (V, E)$ , whose vertex set V represents the set of random variables  $U = \{X_1 \dots X_n\}$ . Directed edges between two nodes  $V_i \rightarrow V_j$  represent a direct dependency between two variables - a missing edge represents the independence of these two variables. The graph to the underlying network is both directed and acyclic, this is abbreviated as DAG [18], [19], [20].

### 4.2 Transformation of a UML Activity

Given a UML activity represented as a graph  $U =$  (activity nodes, activity edges). The set of activity nodes be further divided into sets of nodes:

- $\mathcal{A}$ : Set of actions,
- $\mathcal{S}, \mathcal{E}$ : Initial node and final node,
- $\mathcal{B}$ : Set of decision- and merge nodes (branch nodes),
- $\mathcal{C}$ : Set of fork- and join nodes (concurrency nodes),
- $\mathcal{O}$ : Set of object nodes.

The set of object nodes is given by the set of data pins. A node that is part of one of the node sets  $\mathcal{S}, \mathcal{E}, \mathcal{B}, \mathcal{C}$  is called a control node. Furthermore, the set of activity edges is given by

- $\mathcal{KF}$ : Control flow, i.e. activity edges between actions and control nodes as well as between them underneath each other.
- $\mathcal{DF}$ : Object flow, i.e. activity edges between actions and object nodes or between control nodes and object nodes.

Formally, the translation  $[[U]]$  of a UML activity U to a DAG  $G = (V, E)$  of a Bayesian network is given by:

$$[[(\text{activity nodes, activity edges})]] = (V, E),$$

where:

$$V = \{e_j \mid (e_i, e_j) \in \mathcal{DF}, e_i \in \mathcal{O}, e_j \in \mathcal{A}\} \tag{1}$$

$$\cup \{e_i \mid (e_i, e_j) \in \mathcal{DF}, e_i \in \mathcal{O}, e_j \in \mathcal{A}\} \tag{2}$$

$$\cup \{v \mid v \in \mathcal{A}, \text{Depth}(v) = 1\} \tag{3}$$

$$E = \{(e_j, e_i) \mid (e_i, e_j) \in \mathcal{DF}, e_i \in \mathcal{O}, e_j \in \mathcal{A} : e_j.\text{contains}(\text{“verify”})\} \tag{4}$$

$$\cup \{(e_i, e_j) \mid (e_i, e_j) \in \mathcal{DF}, e_i \in \mathcal{O}, e_j \in \mathcal{A} : e_j.\text{contains}(\text{“calculate”})\} \tag{5}$$

$$\cup \{(e_i, v) \mid v \in \mathcal{A} : \text{Depth}(v) = 1, e_i \in \mathcal{A} : e_i.\text{contains}(\text{“verify”})\}. \tag{6}$$

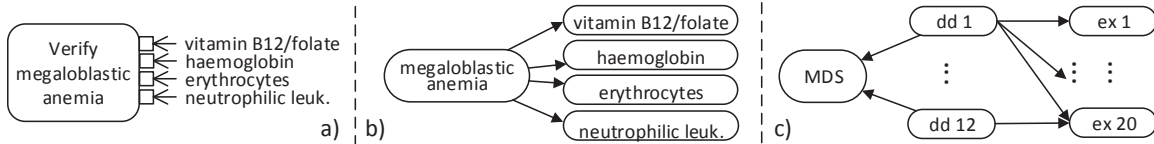


Fig. 5: Subfigures a) and b) depict the translation of a UML action. Node and pins are transformed to vertices in the Bayesian net. There are directed edges from diagnosis to examination values since diseases cause specific test results. Subfigure c) shows the Bayesian structure for a diagnosis of exclusion like MDS.

It is assumed that the nodes of the Bayesian network have a unique name. If a UML activity diagram has several activity nodes with the same name (e.g. as pins of different actions), it is indeed translated several times. But since the destination node has the same name in each case, it is added to the Bayesian network only once. That's because adding an element to a set where the element is already part of, does not alter the set.

UML actions that are provided with input pins representing either a diagnosis, a score or a phase of a disease. Actions are transformed to vertices of the corresponding Bayesian net (1). The pins of these actions are transformed to vertices as well (2). They are providing examination values for an action which can be a non-deterministic verification (diagnosis) or a deterministic calculation (score or phase of a disease due to thresholds etc.).

Since diseases cause typical test results, there is a directed edge from diagnosis to the relevant examination values (4). This is a causal interpretation of an edge<sup>2</sup>. The process of translation is depicted in Subfigures 5a) and 5b). In case of a deterministic calculation the direction of an edge is reversed (5). That is because the examination values are causing a specific score or phase.

The distinction of the cases (4) and (5) are not given by the activity's structure. But the translation can be performed by searching for a specific keyword in the action's name. The function contains checks if a keyword is present. The keyword "verify" represents a non-deterministic verification and "calculate" represents a deterministic calculation.

The translation rules (3) and (6) apply for a diagnosis of exclusion like MDS. In this case we get a special network structure where all differential diagnosis become parents of the diagnosis of exclusion (see Subfigure 5c)).

### 4.3 Network Structure and Parameters

The network structure for a disease like CML follows Subfigure 5b). For an diagnosis of exclusion like MDS, the application of translation rules generate a Bayesian network structure outlined in Subfigure 5c). The conditional probability table (CPT) of MDS is given by

$$P(MDS|D_1, \dots, D_{12}) = \begin{cases} 1, & \text{if } MDS = f(D_1, \dots, D_{12}) \\ 0, & \text{otherwise} \end{cases}$$

$$\text{where: } f(D_1, \dots, D_{12}) = \neg D_1 \wedge \dots \wedge \neg D_{12}. \quad (7)$$

<sup>2</sup>A Bayesian net reflecting the causal structure normally reflects the expert's understanding of a domain and is therefore useful if the parameters of a net are to be assessed by a survey.

Because many differential diagnosis have an impact on several examination values, the corresponding examination nodes can have several parent nodes. For example the examination node haemoglobin has 10 parents. Even if all these involved random variables are binary, this would lead to a haemoglobin CPT with  $2^{10+1}$  values (half of them, i.e.  $2^{10}$  have to be explicitly specified). To build all the CPTs in a robust fashion, we reduced the large amount of values by using some simplifications. We used noisy ORs [21] to reduce the number of parameters for binary nodes with  $n$  parents from  $2^n$  to  $2n$ . Consequently the number of parameters increases linearly with the number of parents. For random variables with more than 2 states the so called noisy MAX model was used to reduce the number of parameters [21], [22], [23]. Due to the lack of a patient database, the network structure was parametrized by the results of a survey. A medical expert was asked to answer 90 questions concerning the MDS diagnostic and to answer 60 questions concerning the CML diagnostic.

### 4.4 Assistance Function

In the models of CML and MDS, the diagnostic node plays a key role. It represents the probability distribution over the states of the disease under consideration. The question is, which evidence can be set to make one of the variable's states as much likely as possible? Regarding a practitioner this means: Which examination value leads to a situation where the decision maker can tell most certain whether the disease is present or not. Consequently, examination values are taken that reduce uncertainty.

A measure of the uncertainty of a random variable is the entropy [24]. It is the higher the more the probability mass scatters over the states of the random variable. In case of a uniform distribution the entropy is maximum. With regard to the diagnostic node that means that the more likely the variable takes one of the 2 states "present" or "not present", the lower the entropy is. Therefore, the reduction of entropy is an essential part of the assistance function.

Let  $A$  be a discrete random variable with  $n$  states  $a_1, \dots, a_n$  and  $P(A)$  the probability distribution. Then the entropy  $\eta$  (in bits) is given by [24]:

$$H(A) = - \sum_{i=1}^n P(a_i) \log_2 P(a_i), \quad (8)$$

$$\text{where: } H(A) \in [0, \log_2(n)].$$

The graph of the entropy for a random variable with 2 states can be described by a concave function. The function reaches its maximum value of  $\log_2(2) = 1$  if both states of

the random variable are equally likely. If one of the 2 states has a probability of 0 the entropy drops to 0, too. In this case we removed the uncertainty about the state of the random variable.

The entropy of a random variable A (e.g. diagnosis) given a random variable B (e.g. examination value) is given by the conditional entropy [24]:

$$H(A|B) = H(A) - I(A, B). \quad (9)$$

$I(A, B)$  is called the mutual information and represents the reduction in the uncertainty of A due to the knowledge of B.

Thus, the benefit of an examination  $E_1$  can be rated by its effect on the uncertainty of a random variable D that represents the (non-)presence of a disease. Hence, formula (9) may be rearranged to give:

$$I(D, E_1) = H(D) - H(D|E_1). \quad (10)$$

If there are several possible examinations  $E_i$  that can be performed, we would prefer the one with the highest mutual information (i.e. highest reduction of uncertainty about D). Thus, a preliminary assistance function recommending an optimal examination  $E^*$  can be obtained:

$$E^* = \arg \max_{E_i} (H(D) - H(D|E_i)). \quad (11)$$

Given a Bayesian net with a node representing the random variable D, the entropy  $H(D)$  can be calculated by formula (8). The conditional entropy  $H(D|E_i)$  is defined as [24]:

$$H(D|E_i) := \sum_{j=1}^n P(E_i = e_j) H(D|E_i = e_j), \quad (12)$$

where  $j=1 \dots n$  are states of random variable  $E_i$ .

The conditioned entropy  $H(D|E_i = e_j)$  in equation (12) is the entropy of the variable D given a certain value  $e_j$  of  $E_i$ . Thus,  $H(D|E_i)$  is calculated by averaging  $H(D|E_i = e_j)$  over all possible values  $e_j$  that  $E_i$  may take.

In a real medical scenario, the choice of an optimal examination solely on basis of equation (11) is not feasible. This would mean that an examination that is very specific for a disease would be suggested first (high benefit), regardless of whether this examination is highly invasive or costly which might outweigh its diagnostic benefit.

The order of the examinations listed in the recommendations of the CPG is a result of the consideration of different interests (costs). Among other things, the invasiveness of an examination is considered. At the beginning of the diagnosis process the proposed examinations are less invasive than at the end of the diagnostic process when the practitioner is more convinced a disease is present or not.

We propose a weighted reduction of uncertainty for integrating cost-benefit considerations during the diagnosis process into our assistance function. The decision maker has the opportunity to choose to what extent to follow the CPG recommendations or to consider a higher reduction of

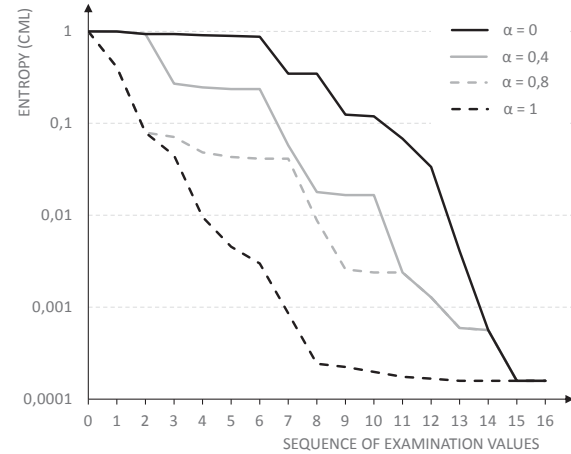


Fig. 6: Entropy of the node CML versus the order of executed examinations. The first examination value obtained is assigned with sequence number 1. Tables 1 lists the assignment of the examination values to the numbers used in the diagram for  $\alpha = 0$ . By following the CPG recommendations, the entropy decreases slowly at the beginning of the diagnostic process (solid black line).

uncertainty during the diagnostic process:

$$\text{Recommendation}_i = (1 - \alpha)(1 - k_i/m) + \alpha I(D, E_i),$$

where  $k_i$ : depth of examination  $E_i$  in CPG ,  
 $m$ : overall depth of CPG ,  
 $\alpha$ : weight factor .

To allow cost-benefit considerations, the assistance function can be adjusted by the parameter alpha. For  $\alpha = 1$  only the reduction of uncertainty is taken into account (benefit). Respectively, for  $\alpha = 0$  the recommendation follows exactly the given CPG model (cost). The increase of the parameter alpha allows to specify how much the practitioner wants to deviate from the CPG recommendations in favor of a uncertainty reduction. The reduction of uncertainty seems to be a realistic reason for not following a CPG recommendation in practical work.

## 5. Verification and Validation

In a first demonstration, the examination values on a patient with suspected CML are assessed in the order given by the corresponding CPG recommendation. The assistance function is therefor parametrized with  $\alpha = 0$  and consequently only the costs given by the CPG are considered.

Figure 6 depicts the reduction of entropy for  $\alpha = 0$  while observing more examination values (solid dark line). By following the CPG recommendations, the entropy decreases slowly at the beginning. With the 11th examination value the entropy has dropped below 0.1. The highest reduction of uncertainty is observed after late examinations. These examinations are highly invasive but are important for identifying the disease for sure. The examination values of the patient are listed in Table 1.

In a second demonstration,  $\alpha = 1$  is chosen. The black dotted line in Figure 6 represents the corresponding progress of the entropy. In this case the assistance function proposes

Table 1: Patient with suspected CML ( $\alpha = 0$ ).

No.	Depth	Examination value	Entropy
0	0	-	1
1	1	Bone pain negative	0,996791632
2	2	Spleen enlarged	0,937185857
3	3	Haemoglobin $< 13g/dL$	0,937185857
4	3	$150k/\mu L \leq$ Thrombocytes $\leq 400k/\mu L$	0,908302337
5	3	$10000/\mu L <$ Leukocytes	0,895383409
6	4	Neutrophilic leukocytes increased	0,873765252
7	4	$1\% \leq$ Basophilic leukocytes $< 20\%$	0,346539286
8	4	Progenitor cells $\leq 30\%$	0,346539286
9	4	$15\% \leq$ Peripheral blasts $< 30\%$	0,124546345
10	5	Progenitor cells bone marrow $\leq 30\%$	0,119056765
11	5	$20\% \leq$ Basophilic leukocytes bone marrow	0,068089434
12	5	$5\% <$ Eosinophilic leukocytes	0,033581937
13	5	Philadelphia chromosome positive	0,004083213
14	5	Bcr-abl positive	0,000563717
15	6	$15\% \leq$ Medullary blasts $< 30\%$	0,000158269
16	6	Lymphatic progenitor cell positive	0,000158269

examinations that maximize the reduction of uncertainty in each step. Highly risky and invasive examinations are recommended first, because they are able to reduce uncertainty the most.

Instead of considering either cost ( $\alpha = 0$ ) or benefit ( $\alpha = 1$ ), our assistance function can be parametrized with  $\alpha \in (0, 1)$ . This allows cost-benefit considerations. For  $\alpha = 0.4$  (solid gray line in Figure 6) the very first examinations (anamnesis, physical examination) are done as the CPG is suggesting. The entropy drops with examination 3. In this examination it is tested whether the so-called Philadelphia chromosome is present or not. The Philadelphia chromosome is a strong indicator for CML [25]. As a result there is an earlier reduction of uncertainty in the diagnostic process.

Figure 7 shows a plot of MDS entropy against the number of a differential diagnosis. The initial value of the entropy is moderately at around 0.5. This is due to the fact that a certain a-priori probability is assigned to each of the 12 differential diagnoses of MDS which influences the probability distribution of MDS. Firstly, the state "present" of MDS is unlikely, even though there is the suspicion of MDS. Medically, this seems plausible and was confirmed by the expert's setting of the corresponding a-priori probabilities of the differential diagnoses: Despite of the suspicion of MDS, first, the more likely differential diagnoses have to be excluded. The more differential diagnoses are excluded, the more likely MDS is present. MDS is present for sure, once all differential diagnoses have been excluded.

The progression of the entropy in Figure 7 reflects this fact. Initially MDS is unlikely, so the plot does not start with the maximum entropy. By successively excluding the differential diagnoses, the entropy first increases to a value close to the maximum of 1. Here  $MDS = positive$  and  $MDS = negative$  are about equally likely. As more and more excluded differential diagnoses make a positive diagnosis of MDS likely, the entropy falls. The individual entropy values for  $\alpha = 0$  is shown in Table 2.

Given is now a patient whose diagnosis has already been established. Patient data suggests that molecular genetics

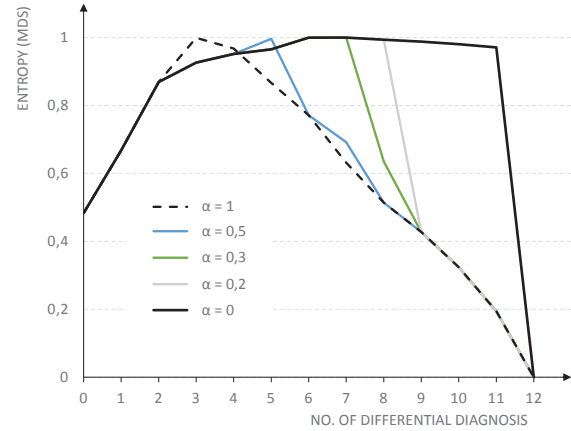


Fig. 7: Entropy of the node MDS versus the order of excluded differential diagnoses for different values of  $\alpha$ . If all differential diagnoses are excluded, MDS is present for sure - the entropy drops to 0.

Table 2: Patient with suspected MDS ( $\alpha = 0$ ).

No.	Depth	Examination	Entropy
0	0	-	0.483273525
1	1	Megaloblastic anemias negative	0.667357598
2	1	Reactive bone marrow alterations negative	0.869280269
3	1	Toxic bone marrow damage negative	0.926423147
4	2	Monocytosis of other etiology negative	0.951739499
5	3	Hypersplenism negative	0.965406878
6	4	Immune thrombocytopenia negative	0.999630334
7	4	Paroxysmal nocturnal hemoglobinuria negative	0.999958433
8	5	Acute leukemias negative	0.993451440
9	6	Congenital dyserythropoietic anemias negative	0.987947563
10	7	Hairy cell leukemia negative	0.980510718
11	7	Aplastic anemia negative	0.970950594
12	7	Myeloproliferative diseases negative	0

and bone marrow histology were actually carried out earlier during the process of diagnosis than is recommended by the guideline. The reproduction of this case in the model yields to a situation in which according to the CPG 4 differential diagnoses still have to be excluded. These are "congenital dyserythropoietic anemia", "hairy cell leukemia", "aplastic anemia" and "myeloproliferative neoplasms" (see Figure 8). The assistance function proposes to exclude "congenital dyserythropoietic anemia" next, since the guideline suggests to do so ( $\alpha = 0$ ). By setting an alpha value of 0.3, it appears clear that the exclusion of myeloproliferative neoplasia in terms of the cost-benefit ratio is better than the exclusion of congenital dyserythropoietic anemia. This is exactly what happened in the real world: the practitioner recognized the higher probability of a "myeloproliferative neoplasia" because of his expertise, and decided to exclude this particular disease by molecular genetics and bone marrow histology.

To sum up, it can be said that the models for CML and MDS behave as expected. They are able to describe the corresponding diagnostic process and can provide the intended assistance function. Real world cases can be reproduced with appropriate alpha values. The parameter alpha ranges somewhere in between 0 and 1 (avoiding the extreme values 0 and 1) since medical cases are based on the consideration



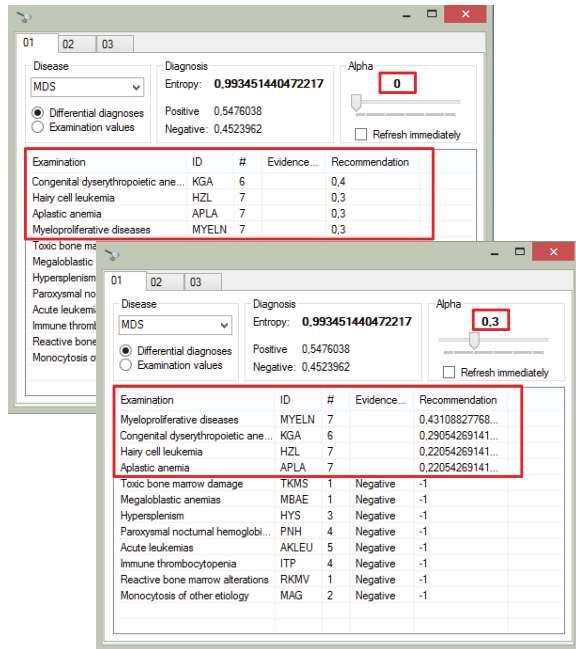


Fig. 8: User interface of the assistance function. There are 4 differential diagnoses left.

of the CPG recommendations as well as the reduction of uncertainty about the presence of a disease.

## 6. Conclusion

In this paper we introduced the concept of an interface between experts and CPG models used for supporting the practitioner during the diagnostic process. Our approach utilizes UML activities as a basis for a CPG formalization done by a medical and/or technical expert. An activity can then be translated into different CPG models such as Petri nets or Bayesian nets. These models are used to provide the actual assistance function. In this work we focused on the transformation of UML activities into Bayesian networks.

The assistance function provided by a Bayesian net is based on the reduction of uncertainty and the sequence of recommended examinations given by a CPG. During the diagnostic process it provides the practitioner with recommendations which next examination value to take. In contrast to assistance functions based on rigid CPG models, our assistance function allows to weigh up costs and benefits of a certain examination. To which degree the reduction of uncertainty outweighs the costs (e.g. due to invasiveness of an examination) can be adapted individually.

For the future, we plan to take our approach to the next level by integrating it in a real world diagnostic application that helps to overcome barriers in CPG implementation and reduces the gap between theoretical medical knowledge and practical solutions.

## References

[1] National Center for Biotechnology Information, U.S. National Library of Medicine, "Pubmed," March 2015, Accessed: 2015/03/23. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed>

[2] C. D. Mulrow, "Rationale for systematic reviews." *BMJ: British Medical Journal*, vol. 309, no. 6954, p. 597, 1994.

[3] A. d. W. M. F. AWMF, "Das Leitlinien-Manual von AWMF und ÄZQ," *Zentrum ärztliche Qualitätssicherung*, vol. 95 (Suppl I), 2001.

[4] E. Field and K. Lohr, *Guidelines for Clinical Practice: From Development to Use*. National Academies Press, 1992.

[5] E. Steinberg, S. Greenfield, M. Mancher, et al., *Clinical Practice Guidelines We Can Trust*. National Academies Press, 2011.

[6] J. Grimshaw and I. T. Russell, "Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations," *The Lancet*, vol. 342, no. 8883, pp. 1317–1322, 1993.

[7] L. A. Bero, R. Grilli, J. M. Grimshaw, E. Harvey, A. D. Oxman, and M. A. Thomson, "Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings," *Bmj*, vol. 317, no. 7156, pp. 465–468, 1998.

[8] P. Alonso, A. Irfan, I. Solà, I. Gich, M. Delgado Noguera, D. Rigau, S. Tort, X. Bonfill, J. Burgers, and H. Schunemann, "The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies," *Quality and Safety in Health Care*, vol. 19, no. 6, pp. 1–7, 2010.

[9] A. Giguère, F. Légaré, J. Grimshaw, S. Turcotte, M. Fiander, A. Grudniewicz, S. Makosso Kallyth, F. M. Wolf, A. P. Farmer, and M.-P. Gagnon, "Printed educational materials: effects on professional practice and healthcare outcomes," *Cochrane Database Syst Rev*, vol. 10, 2012.

[10] H. Kirchner, M. Fiene, and G. Ollenschläger, "Bewertung und Implementierung von Leitlinien," *Die Rehabilitation*, vol. 42, no. 2, pp. 74–82, 2003.

[11] H. Störrle, "Semantics of control-flow in UML 2.0 activities," in *Visual Languages and Human Centric Computing, 2004 IEEE Symposium on*, 2004, pp. 235–242.

[12] H. Störrle, "Semantics and verification of data flow in UML 2.0 activities," *Electronic Notes in Theoretical Computer Science*, vol. 127, no. 4, pp. 35–52, 2005.

[13] C. Kecher, *UML 2: das umfassende Handbuch*. Bonn: Galileo Press, 2011.

[14] C. Rupp, S. Queins, et al., *UML 2 glasklar*. München: Carl Hanser Verlag GmbH Co KG, 2012.

[15] OMG, "OMG Unified Modeling Language(OMG UML) Superstructure Version 2.4.1," 2011, Accessed: 2014/08/28. [Online]. Available: <http://www.omg.org/spec/UML/2.4.1/Superstructure/PDF/>

[16] H. Störrle, *UML 2 für Studenten*. Pearson Studium, 2005, vol. 320.

[17] K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*. cRc Press, 2003.

[18] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[19] R. E. Neapolitan, *Probabilistic reasoning in expert systems: theory and algorithms*. Wiley, 189.

[20] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[21] M. Henrion, "Practical issues in constructing a Bayes belief network," *Uncertainty in Artificial Intelligence*, vol. 3, pp. 132–139, 1988.

[22] F. J. Diez, "Parameter adjustment in Bayes networks. The generalized noisy OR-gate," in *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, 1993, pp. 99–105.

[23] S. Srinivas, "A generalization of the noisy-or model," in *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, 1993, pp. 208–215.

[24] T. M. Cover and J. A. Thomas, *Elements of information theory*. Hoboken, New Jersey: John Wiley & Sons, 2006.

[25] A. Hochaus, G. Baerlocher, T. H. Brümmendorf, Y. Chalandon, P. le Coutre, G. Dölken, C. Thiede, and D. Wolf, *Chronische Myeloische Leukämie (CML) Leitlinie - Empfehlungen der Fachgesellschaften zur Diagnostik und Therapie hämatologischer und onkologischer Erkrankungen*. Berlin: DGHO Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e.V., 2013.

# ECG Analysis for Automated Diagnosis of Subclasses of Supraventricular Arrhythmia

Purva R Gawde<sup>1</sup>, Arvind K. Bansal<sup>2\*</sup>, and Jeffrey A. Nielson<sup>3</sup>

<sup>1</sup>Department of Computer Science, Kent State University, Kent, OH, USA, {pgawde@kent.edu}

<sup>2</sup>Department of Computer Science, Kent State University, Kent, OH, USA, {akbansal@kent.edu}

<sup>3</sup>Emergency Division, Summa Health Systems, Akron, OH, USA, {jeffnielson@gmail.com}

\*Corresponding author

**Abstract** – A Markov-model based technique is proposed to automatically detect a patient's disease state into different subclasses of Supraventricular arrhythmia based upon automated ECG analysis. Separate Markov-models have been developed for each subclass using a finer resolution of P-waves that takes into account left and right atria and multiple slopes of the P-waves. ECG data from Physionet database has been used to train the Markov-models. Patient's ECG has been transformed to a probabilistic-transition-graph. Graph based comparison has been used to match probabilistic-transition-graph derived from Patient's ECG and Markov-models of the corresponding subclasses to identify the patient's disease state in real time. The result correlates well with the physician's diagnosis of supraventricular arrhythmia. Algorithms and sensitivity analysis have been presented.

**Keywords:** Arrhythmia, ECG analysis, Markov model, Patient diagnosis, Patient monitoring

## 1 Introduction

As society is aging, the mortality caused by cardiovascular diseases is increasing [4]. One of the major problems is sudden death caused by arrhythmias [4, 19, 22]. During arrhythmias, heart-beat becomes rapid, irregular, and unsynchronized. Statistics from the Center for Disease Control and Prevention (CDC) estimates sudden cardiac death rates due to arrhythmias at more than 600,000 per year [4]. Arrhythmias occur suddenly with varying refractory period, may disappear, and may not be able to reproduce during regular checkup [12]. Recording and classification of the subclasses of arrhythmias when it occurs can timely save many lives for our aging yet mobile society.

Arrhythmia is broadly classified as: 1) supraventricular arrhythmia [9, 12, 14, 22] that originate due to electric impulse irregularity and atria enlargement in the upper chambers; and 2) ventricular arrhythmia [12, 21] that originate due to conduction irregularity in the ventricles. Arrhythmia classification is necessary for attending physicians after cardiac surgery as some types of arrhythmias such as supraventricular arrhythmia [9, 22] occurs frequently post-cardiac surgery [19, 20, 22], and can be fatal if not treated immediately [19, 20, 22].

Depending on the origin of electric impulses, supraventricular arrhythmia is classified into: *atrial fibrillation (AFib)*, *atrial flutter (AFlu)*, *junctional tachycardia (JTachy)*, *atrial-ventricular nodal reentry tachycardia (AVNRT)* and *ectopic atrial tachycardia (EAT)*.

Previous approaches for automated detection of subclasses of arrhythmias [8, 13, 17, 26] have met with limited success due to the lack of electric pulse synchronization that causes inconsistent morphological changes in P-QRS-T waveforms and their occurrence-sequence due to irregular beats. Previous works are limited to the separation of supraventricular arrhythmia from ventricular arrhythmia [2, 8, 26] and limited identification of a subclasses of arrhythmias [7, 10, 13, 17, 21].

Different subclasses of supraventricular arrhythmia are treated differently [9, 14, 22] using different medications [14, 20, 22]. Misdiagnosis of different forms of supraventricular arrhythmia can be fatal [19, 22] or mistreatment can cause serious side-effects [10, 19, 20, 22]. Hence, correct automated diagnosis of subclasses of supraventricular arrhythmia is important.

This paper describes an improvement in automated diagnosis of a patient suffering from supraventricular arrhythmia. We develop Markov-models for major subclasses of supraventricular arrhythmia, and employ these models for the diagnosis of a patient's condition in real time.

The major technical contributions of this paper are:

1. Resolution of P-waves into four states in a Markov-model based on electric impulse traversal in left and right atria and the positive and negative slopes of P-waves. This resolution generates distinct Markov-models for each subclass of supraventricular arrhythmia.
2. Construction of distinct Markov-model for each subclass of supraventricular arrhythmia by studying Physionet public database [23]; and
3. Modeling patient's ECG as probabilistic-transition-graph (PTG), and matching PTG with the probability-graphs corresponding to the Markov-models to identify exact disease states.

The paper is organized as follows: Section 2 describes the background information about subclasses of supraventricular

arrhythmia and needed definitions related to graph matching. Section 3 describes our approach and the training of the Markov-models with finer classification of P-waves for each subclass of supraventricular arrhythmia. Section 4 describes an algorithm to identify disease state using probabilistic-graph matching. Section 5 describes an implementation. Section 6 discusses the results. Section 7 compares this work with other related works. The last section concludes the work.

## 2 Background

*Supraventricular arrhythmia* is a rapid heart rhythm where electrical impulse travels from the upper chambers (SA-node or ectopic foci in atria) to ventricles via AV-node [9, 12]. Supraventricular arrhythmia occurs due to the lack of synchronization of depolarization and repolarization caused by rapid misfiring (or lack of firing) of SA-node, presence of a cyclic electric current in atria, and/or impulse arising from junction valves or one or more ectopic mass in atria. There are five major subclasses, namely, *atrial fibrillation (Afib)*, *atrial flutter (AFlu)*, *junctional tachycardia (JTachy)*, *atrial-ventricular nodal reentry tachycardia (AVNRT)*, and *Ectopic atrial tachycardia (EAT)*.

### 2.1 Supraventricular arrhythmia subclasses

*Atrial fibrillation (Afib)* occurs when action potentials fire very rapidly within the pulmonary veins or atria in a chaotic manner [7, 12, 17]. The result is a very fast atrial rate (about 400-600 beats per minute). The AV-node becomes intermittently refractory and allows only a fraction of atrial action potentials to reach the ventricles.

*Atrial flutter (AFlu)* occurs when a *reentrant electrical circuit* is present in atria causing a repeated loop of electrical activity to depolarize the atrium at a rate of about 250-350 beats per minute [12, 17, 22]. This produces a characteristic "sawtooth" pattern. Only a fraction of P-waves are conducted through the AV-node. Typical *Aflu* rotates "counterclockwise" in direction, and results in negatively directed flutter waves in the inferior leads.

*Junctional tachycardia (JTachy)* arises from the area encompassing AV-node, his-bundle and immediately surrounding atrial tissue [12]. Depolarization proceeds towards atria. Hence P-waves in inferior leads are negative. Discharge has very little distance to travel to the AV-node. Hence, PR interval is very short. In some cases, atrial depolarization coincides with ventricular depolarization. In that case, P-wave gets buried in QRS-complex. It is also possible that atrial depolarization follows ventricular depolarization. Then P-wave occurs after QRS-complex.

*Atrial-ventricular nodal reentry tachycardia (AVNRT)* is a common form of nodal arrhythmia [9, 12, 14]. In *AVNRT*, atrial depolarization is conducted through AV-node via two tracts instead of one [9, 12, 14]. One tract functions normally and other conducts the impulse slowly. An impulse travels over the slow pathway towards the ventricles and returns via the faster pathway to the atria. The retrograde P-wave is

embedded inside QRS-complex or appears after the QRS-complex.

*Ectopic atrial tachycardia (EAT)* has discrete foci in the main mass of atria outside SA-node and junctional region [9, 12]. Depending on the origin of impulse atrial depolarization may (or may not) travel from right to left atria. Since the impulses travel from different ectopic foci in atria, P-wave axis keeps changing. In this case, P-wave is negative in one of the following leads I, II, III or V5 or V6.

### 2.2 Definitions

A *weighted-graph* has weighted-edges. An *unweighted-edge* is denoted as *source-node*  $\rightarrow$  *destination-node*. A *weighted-edge* is denoted as a pair of the form (*source-node*  $\rightarrow$  *destination-node*, *weight*). A *probability-graph* is a weighted-graph with probability as edge-weights. *Markov-model* [24] is a probabilistic finite-state automata that is modeled as a probability-graph that has one or more initial and final states. The edges in the probability-graph for a Markov-model shows the transition-probability from one state to another. A probability-graph is represented using a transition matrix such that the value in a cell (i, j) represents the transition-probability from the state *i* to state *j*; and row *i* ( $1 \leq i \leq$  number of nodes) represents the weighted outgoing edges to other nodes. The column-id of cell with the highest value in a row gives the next most-probable transition state. The *probability of a path* in the probability-graph is given as the product of the edge-weights in the path.

*Probabilistic-transition-graph*, denoted by PTG, is a probability-graph that is constructed by analyzing a patient's ECG for a finite duration of real-time monitoring.

*Most-probable-path*, denoted by MPP, is the path with highest edge-weights from the initial state to the final state in a probability-graph. The edge-weights of MPP are maximum of the outgoing-edges from the corresponding source-nodes of the edges. The probability of an MPP in PTG is closest to the maximum value of 1.0 when compared to other paths.

*Probability-difference*, denoted as  $\delta_i$  ( $1 \leq i \leq$  number of edges in a path), is the absolute-difference  $|\beta_{1i} - \beta_{2i}|$  where  $\beta_{1i}$  and  $\beta_{2i}$  are edge-weights of the corresponding edges in the two paths of probability-graphs. *Path-difference* is defined as the sum of probability-differences ( $\sum \delta_i$ ) of the corresponding paths in two probability-graphs. Path-difference acts as a metrics of dissimilarity for the comparison of MPP and the corresponding path of a probability-graph: lower path-difference means that PTG is more similar to the probability-graph. It can be mathematically derived that path-difference between MPP in PTG and the corresponding path in a probability-graph is a close approximation of the absolute-difference of transition-probabilities  $|\prod \beta_{1i} - \prod \beta_{2i}|$  ( $1 \leq i \leq$  number of edges in the MPP) since  $\delta_i \ll \beta_{1i}$  and  $\delta_i \ll \beta_{2i}$  in the MPP, and subterms containing factors  $\delta_i \delta_j$  can be dropped from the product expansion during path-comparison as they have very small value. Similarly, the factors  $\beta_i \delta_j$  is limited by the value  $\delta_j$  since  $\beta_i$  is less than and close to 1.0.

### 3 Approach

The overall approach is: 1) develop a Markov-model that captures all the variations of electric-impulse travel in upper chambers; 2) train the Markov-models for each subclass of supraventricular arrhythmia using well-annotated public database such as Physionet [23] that contains large number of cases of supraventricular arrhythmia; 3) perform real-time analysis of patient's ECG to get a probabilistic-transition-graph (PTG); 4) identify the most-probable-path in a PTG; and 5) develop graph matching algorithms to identify the best matching probability-graph corresponding to subclasses of supraventricular arrhythmia and the PTG. Best match is identified by: 1) deriving the most-probable-path (MPP) in PTG; 2) identifying that PTG is a subgraph of probability-graphs corresponding to subclasses of supraventricular arrhythmia; 3) pairwise comparing MPP of PTG with probability-graphs filtered from step 1, one at a time, to derive path-differences between MPPs and the corresponding paths in the probability graphs 4) sorting path-differences to identify Markov-model with minimum path-difference.

Since P-waves are related to the travel of electric impulse in upper chambers of the heart, we exploited various possibilities of P-waves in the development of Markov-models for the subclasses of supraventricular arrhythmia.

#### 3.1 Markov-model

Our proposed Markov-model separates P-waves into 4 different states to model: 1) split of P-waves from left and right atria; and 2) the slope of the P-waves. P-waves are split into four different segments: (1) P-wave rising-edge of the right atrium denoted as  $P_{11}$ ; (2) P-wave falling-edge from the right atrium denoted as  $P_{12}$ ; (3) P-wave rising-edge from the left-atrium denoted as  $P_{21}$ ; and (4) P-wave falling-edge from the left atrium denoted as  $P_{22}$ . In the case of P-wave inversion, the falling-edge is labeled  $P_{12}$ , and the following rising-edge is called  $P_{21}$ .

There are three possibilities of P-wave splitting: (1) *Case 1* - P-wave from right atrium is almost fully superimposed with P-wave from left-atrium; (2) *Case 2* - falling edge of P-wave from right atrium is partially superimposed with rising edge of the P-wave from left atrium; and (3) *Case 3* - P-wave from right atrium is completely separated from P-wave from left atrium. The three cases of P-waves splitting are modeled as: (1)  $P_{11} \rightarrow P_{22}$  for the case 1 when both P-waves are superimposed; (2)  $P_{11} \rightarrow P_{12} \rightarrow P_{21} \rightarrow P_{22}$  when P-waves from two atria are partially or fully split (cases 2 and 3).

The proposed Markov-model has 11 states:  $ISO_3 - P_{11} - P_{12} - P_{21} - P_{22} - ISO_1 - Q - R - S - ISO_2 - T$  where  $ISO_1$  represents the baseline for PQ-segment,  $ISO_2$  represents the baseline for a ST-segment, and  $ISO_3$  represents baseline for the segment between a T-waveform and the next  $P_{11}$ . The state  $ISO_3$  is also the initial state as well as final state. Modeling isoelectric lines between waveforms as separate states is important to capture delays in the depolarization and repolarization cycle, and is needed to reason about missing or misplaced P-waves, S-waves and T-waves due to the lack of synchronization in arrhythmia.

A Markov-model is obtained using the annotations obtained through wave-peak detections ( $P_{11}$ ,  $P_{12}$ ,  $P_{21}$ ,  $P_{22}$ , Q, R, S and T), wave-onsets and wave-offsets ( $ISO_1$ ,  $ISO_2$ ). For some cases, observed waveform could be either T-wave (in case of missing ST-segment) or elevated ST-segment (T-wave superimposed with ST-segment). Those waveforms are marked as ST/T-node in the Markov-model (see Figures 1.a and 1.b).

#### 3.2 Trained Markov-models

Markov-models were trained using the Supraventricular Arrhythmia Database available at Physionet [23] using approximately 15 cases for each Markov-model. Approximately 1800 heartbeats were used for training each Markov-model.

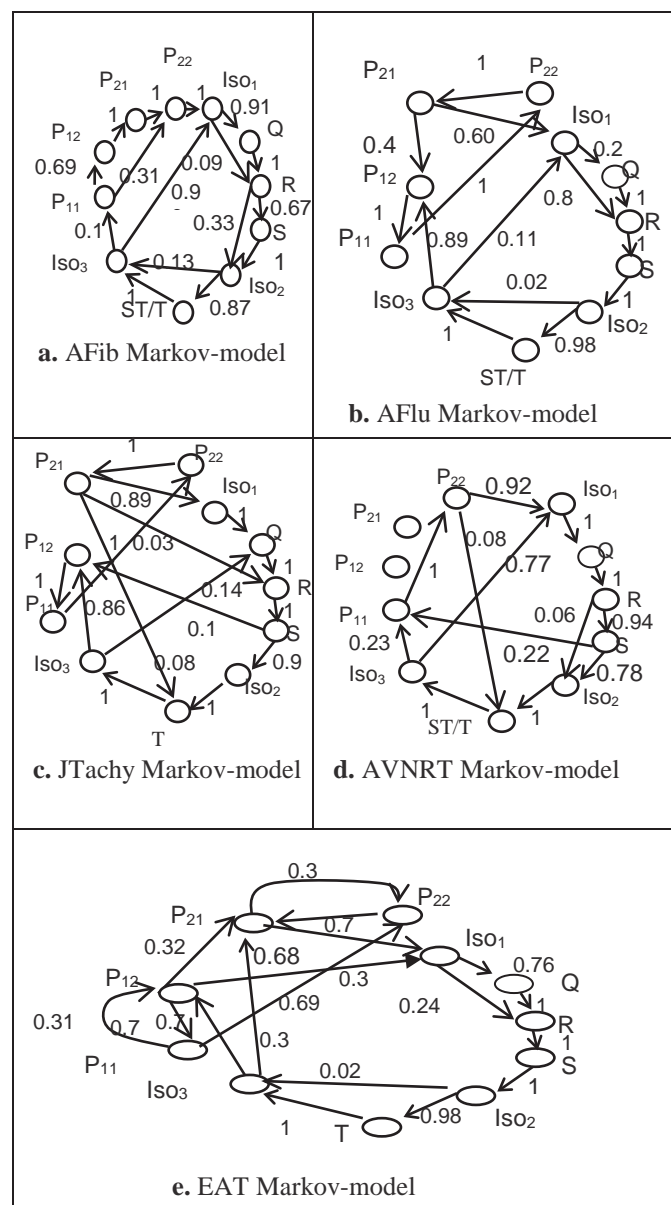


Figure 1. Trained Markov-models

Figure 1.a describes a trained Markov-model for *atrial fibrillation (AFib)*. The transitions are:  $Iso_3 \rightarrow P_{11}$  (P-wave present, probability 0.1);  $Iso_3 \rightarrow Iso_1$  (P-wave missing, probability 0.9);  $P_{11} \rightarrow P_{22}$  (superimposed P-waves, probability 0.31);  $P_{11} \rightarrow P_{12} \rightarrow P_{21} \rightarrow P_{22}$  (P-wave splitting, probability 0.69);  $P_{22} \rightarrow Iso_1$  (always when P-wave occurs);  $Iso_1 \rightarrow Q$  (Q-wave present, probability 0.91);  $Iso_1 \rightarrow R$  (Q-wave missing, probability 0.09);  $Q \rightarrow R$  (probability 1.0),  $R \rightarrow S \rightarrow Iso_2$  (S-wave present, probability 0.67);  $R \rightarrow Iso_2$  (S-wave missing, probability 0.33);  $Iso_2 \rightarrow T \rightarrow Iso_3$  (T-wave present, probability 0.87); and  $Iso_2 \rightarrow Iso_3$  (T-wave missing, probability 0.13). Multiple sites within atria depolarize independently. Low amplitude action-potentials fire rapidly resulting into missed P-waves as shown by the transition  $Iso_3 \rightarrow Iso_1$ .

Figure 1.b describes a trained Markov-model for *atrial flutter (AFlu)*. The discharge of depolarizing current from the re-entrant loop may produce large negative deflections in the inferior leads causing negative P-waves. P-waves are missed occasionally due to fast atrial activity. Negative P-waves are shown by the transition-sequence  $P_{12} \rightarrow P_{11} \rightarrow P_{22} \rightarrow P_{21}$ . The transition  $P_{21} \rightarrow P_{12}$  is due to the possibility of more than one P-wave for one QRS-complex. Sometimes T-wave gets superimposed with the next P-wave that shows up as missing T-wave as illustrated by the direct transition  $Iso_2 \rightarrow Iso_3$ .

Figure 1.c describes a trained Markov-model for *junctional tachycardia (JTachy)*. The probability transition-sequence  $P_{12} \rightarrow P_{11} \rightarrow P_{22} \rightarrow P_{21}$  is due to the presence of negative P-waves in the inferior leads. The P-waves occasionally coincide with the QRS-complexes as illustrated by transition  $Iso_3 \rightarrow Q$ . In some cases, P-wave appears after QRS-complex as illustrated by the transition from  $S \rightarrow P_{12}$ .

Figure 1.d describes a trained Markov-model for *AV nodal reentrant tachycardia (AVNRT)*. The transition  $Iso_3 \rightarrow Iso_1$  illustrates retrograde P-wave due to the superposition with QRS-complex. The transition  $S \rightarrow P_{11}$  shows P-wave appearing after QRS-complex. T-wave appearing after P-wave is shown by the transition  $P_{22} \rightarrow T$ .

Figure 1.e describes a trained Markov-model for *ectopic atrial tachycardia (EAT)*. Transition from  $Iso_3$  may go to  $P_{11}$  (for positive P-wave) or to  $P_{12}$  (negative P-wave). Since origin of impulse is not SA-node, impulse traveling from AV-node is weak causing undetectable low amplitude Q-waves. Hence, the transition  $Iso_1 \rightarrow R$  occurs with sufficient probability (0.24).

## 4 Algorithm to identify patient's state

Probabilistic-transition-graph (PTG) for patients are constructed from ECG waves in real-time using a moving window of 30 seconds to monitor a patient in real-time. The problems of identifying patient's disease state reduces to matching the PTG with the probability-graphs corresponding to Markov-model of each arrhythmia-subclass and identifying the closest match.

The algorithm to match the PTG with the probability-graphs has four steps:

**Step 1:** identifying the subset of probability-graphs corresponding to the Markov-models of arrhythmia-

subclasses that contain all the edges of PTG. If any edge in PTG is missing from a probability-graph, then the corresponding arrhythmia-subclass is ruled out;

**Step 2:** deriving the most-probable-path (MPP) in PTG;

**Step 3:** pairwise matching the MPP in PTG with the subset of probability-graphs derived in Step 1;

**Step 4:** sorting the path-differences of the matched paths to identify the probability-graph with the smallest path-difference.

Path-difference between MPP in PTG and the corresponding path in a probability-graph is calculated by iteratively summing up the probability-difference  $\delta_i$  (difference in the edge-weights) between the corresponding edges in two paths.

An algorithm is described in Figure 2. The probability-graphs are represented as transition-matrices denoted by  $T_i$  ( $1 \leq i \leq$  number of transition-matrices). The cell values in  $T_i$  represent the transition-probabilities. PTG is modeled as a transition-matrix denoted by  $T^P$ . The cell-value in the  $i$ th row and  $j$ th column of  $T^P$  is denoted by  $T^P(i, j)$ , and shows the weight of the edge  $i \rightarrow j$ . A row in a transition-matrix contains the edge-weights of outgoing-edges from a node; and the highest cell-value in a row gives the next node in the MPP.

The set of probability-graphs that contain PTG as a subgraph are denoted by  $S^M$ . PTG is identified as a subgraph of one or more probability-graphs by testing that the set of unweighted-edges  $S^P$  in the transition-matrix  $T^P$  is a subset of the set of unweighted-edges  $S_i$  in the transition-matrices  $T_i$  ( $1 \leq i \leq$  number of probability-graphs). If there exists an edge  $E \in S^P$  such that  $E \notin S_i$  then the corresponding  $T_i$  is not included in the set  $S^M$ .

The pseudo-code to derive MPP is based upon: 1) starting from the initial state  $Iso_3$ ; and 2) iteratively finding the outgoing weighted-edge with highest weight ( $cur\_state \rightarrow next\_state, max\_prob$ ) and including the weighted-edge in the accumulator  $ptg\_mpp$  until the  $Iso_3$  (final state) is reached again.  $Iso_3$  is both initial and final state since P-QRS-T cycle repeats. The variable  $ptg\_mpp$  accumulates the sequence of edges in the MPP.

To match the MPP in PTG and the corresponding path in a probability-graph, the variable  $ptg\_mpp$ , is copied in the variable  $edge\_set$ . The edges in  $edge\_set$  are matched with the edges in edge-set  $S_i$  of  $T_i$  using an iterative-loop. Next edge, denoted by the variable  $cur\_edge$ , is extracted from the  $edge\_set$ , and the corresponding outgoing-edge  $E$  is identified in the edge-set  $S_i$ . The absolute-difference of the weights of the two edges  $|w_1 - w_2|$  is added to an accumulator  $path\_diff$ . After the variable  $edge\_set$  becomes empty, all the edges in the MPP have been compared, and the variable  $path\_diff$  contains the path-difference between two paths. The pair  $(path\_diff, D_i)$  is inserted into a set  $matched\_set$  where  $D_i$  is the name of the corresponding disease-state.

After pairwise comparing MPP and the corresponding paths of the probability-graphs, the elements on the set  $matched\_set$  are sorted in the ascending order of path-difference values. The corresponding disease state is identified as the second field of the first element in the sorted-sequence.

**Algorithm** Diagnose-patient-state

**Input:** 1. Transition-matrix  $T^P$  for the patient's PTG;  
 2. Set of transition-matrices  $\{T_1, \dots, T_M\}$  for the probability-graphs;  
 3. set of the corresponding disease-states  $\{D_1, \dots, D_M\}$ ;  
**Output:** Diagnosed disease-state  $D^F$ ;

```
{ % Find the probability-graphs that include PTG
  SP = set of unweighted-edges in the transition-matrix TP; SM = { };
  for each Ti (1 ≤ i ≤ M) do { % identify probability-graphs
    Si = set of unweighted-edges in Ti;
    if (SP ⊆ Si) SM = SM + Ti; }

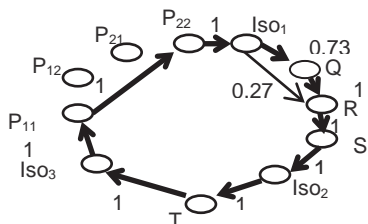
% Find the most-probable-path in the transition-matrix TP;
initial_state = 'Iso3'; final_state = 'Iso3';
cur_state = initial_state; ptg_mpp = { };
do { max_prob = maximum(TP(cur_state, col) for 1 ≤ col ≤ 13);
    next_state = col such that TP(cur_state, col) = max_prob;
    ptg_mpp = ptg_mpp + (cur_state → next_state, max_prob);
    cur_state = next_state; }
until (cur_state == final_state);

% Match the MPP in PTG with the probability-graphs
matched_set = { }; % initialize matched_set to an empty set
for each Ti ∈ SM do { % find matching path in probability-graph
  edge_set = ptg_mpp; path_diff = 0.0;
  Si = set of weighted-edges in Ti;
  while (non_empty(edge_set)) {
    cur_edge = next_element(edge_set); % get next edge
    let cur_edge be of the form (src → dest, w1);
    search the edge E of the form (src → dest, w2) ∈ Si;
    path_diff = path_diff + absolute_value(w1 - w2);
    edge_set = edge_set - cur_edge; }
  matched_set = matched_set + (path_diff, Di); }
sorted_set = ascending_sort(matched_set);
DF = second_field(first_element(sorted_set));
return(DF); }
```

**Figure 2.** Matching patient's PTG with Markov-models.

**Example 1**

The ECG of a patient was procured from the MIT Physionet database [23]. Patient is a 36-year-old female with a heart rate of 270 bpm (beats per minute). She had a prior history of *atrial fibrillation*, and was suffering from *right atrial hypertrophy*. The PTG is shown in Figure 3. The path with bold edges shows the MPP  $Iso_3 \rightarrow P_{11} \rightarrow P_{22} \rightarrow Iso_1 \rightarrow Q \rightarrow R \rightarrow S \rightarrow Iso_2 \rightarrow T \rightarrow Iso_3$ .



**Figure 3.** Probability-transition-graph for a patient

The edge  $P_{22} \rightarrow Iso_1$  is missing from the probability-graphs of *AFlu*, *JTachy* and *EAT*. The edge  $P_{11} \rightarrow P_{22}$  is

missing from the probability-graph of *AVNRT*. Hence, the arrhythmia-subclasses *AFlu*, *JTachy*, *EAT*, and *AVNRT* are ruled out. All the edges of PTG are present only in the probability-graph of *AFlu*. Hence, the patient's disease state is identified as *atrial fibrillation* that matches with the clinical interpretation. The overall path-difference between the MPP of PTG and the corresponding path in *AFlu* is 1.42.

## 5 Implementation

The software implementation consists of five interconnected modules: 1) *module-1*: P-QRS-T-waveform detection and annotation; 2) *module-2*: slope detection and annotation of  $P_{11}$ ,  $P_{12}$ ,  $P_{21}$  and  $P_{22}$ ; 3) *module-3*: development of transition matrix for Markov-model from P-QRS-T annotation; 4) *module-4*: development of transition matrix for analyzing patient's ECG; and 5) *module-5*: graph matching algorithm. The software used available C-library functions in WFDB (Waveform Database) library [28] (available at <http://www.physionet.org/physiotools/wfdb.shtml#download>) for module 1. The output of this software gave us the onset, offset and amplitude of P-waves, QRS-complex and T-waves. Remaining modules were developed in C-language to be consistent with WFDB library [28].

The output from the module 1 and the ECG data-file were used as input to the module-2. Module-2 derived slopes of P-waves; split QRS-complex into Q-slope, R-slope, and S-slope; and split P-wave into four segments:  $P_{11}$ ,  $P_{12}$ ,  $P_{21}$ , and  $P_{22}$ . P-wave split was identified using differentiation of the waveform and comparing the point(s) of zero-slope (slope = 0) with the offset-point of P-wave. The presence of two zero-slope (slope = 0) points describes P-wave split into P-wave from the right atrium followed by P-wave from the left atrium. In the case of P-wave split, first rising edge was called  $P_{11}$ , followed by first falling edge as  $P_{12}$ , followed by the second rising edge as  $P_{12}$ , followed by the second falling edge as  $P_{22}$ . In the absence of second zero-slope point, the P-waves from left and right atrium were superimposed; the first rising edge was labeled as  $P_{11}$ , and the following falling edge was labeled  $P_{22}$ . In the case of P-wave inversion, shown by first edge as falling edge, first falling edge was labeled as  $P_{12}$ , and the following rising edge was labeled as  $P_{21}$ .

The observations (amplitudes and durations of waveforms) from ECG signal were used to derive transition probabilities for each state. Markov-model was developed for each subclass by obtaining transition probabilities using Baum-Welch Algorithm [24]. Baum-Welch algorithm is an expectation-maximization technique. Using this technique, those edges with very small probability below a threshold (threshold-probability < .001) were discarded to reduce noise. The graph-matching algorithm was coded in C-language.

### 5.1 Database

The data for training the Markov-model was obtained from MIT Physionet database [23]. The database contains 48 fully annotated half-hour excerpts of two-channel ambulatory ECG recordings. The recordings were digitized at 360 samples per second per channel.

## 6 Results and discussions

In this section, we compare the statistics based upon our analysis of the patients' information available in a public database available at Physionet [23] and St. Petersburg 12-lead database [25]. The data was extracted and analyzed by an expert cardiologist to identify various subclasses of supraventricular arrhythmia. The data of 30 patients were automatically analyzed using multiple windows of 30 seconds to generate PTGs of the patient's condition. The ECG in the windows was reconfirmed for the correctness by the physician, and was automatically analyzed using our graph matching algorithm. The result of sensitivity analysis (Sensitivity = true positives / (true positives + false negatives)) is summarized in Table I.

**Table I.** Statistics of ECG diagnosis

Disease State	Number of Patients	Sensitivity
<i>AFib</i>	30	92%
<i>AFlu</i>	30	89%
<i>AVNRT</i>	30	90%
<i>JTachy</i>	30	84%
<i>EAT</i>	30	81%

Table I shows that the subclasses of supraventricular arrhythmia are identified with high percentage of accuracy. *JTachy* and *EAT* are difficult to identify due to occasional and asynchronous superimposition of P-waves in QRS waveforms or T-waves. Few misclassification of *AFib* into *AFlu* or *AVNRT* are due to the close probability of QRS-complex in the case of missing P-wave. Misclassification of *AFlu* into: 1) *EAT* are due to negative P-waves; and 2) *AFib* are due to similar transition probability of QRS-complex in case of missing P-waves. Misclassifications of *AVNRT* into *AFib* are due to similar transition probability of QRS-complex. Misclassification of *JTachy* into *Aflu* or *EAT* are due to the negative P-waves. Misclassifications of *EAT* into *AFlu* are due to the negative P-waves and missing Q-waves. Misclassifications can be reduced by performing statistical analysis on multiple PTGs derived from the same patient and increasing the size of the training database.

The overall analysis time varied between 60 – 100 milliseconds based upon the number of ECG waveforms present in 30 second window. Due to the varying frequency of waveforms present in different patients at different times, the analysis time was different for different cases. However, the analysis is done in real-time.

## 7 Related works

ECG analysis has been done for wave detection, wave segmentation and arrhythmia classification using various techniques including Hidden Markov-model [2, 6, 8, 26], Wavelet transform [13, 17], and combination of Hidden Markov-model and wavelet transform [3].

Some algorithms have been developed based on wavelet transform [5, 13, 18] for detecting ECG characteristic points

that can distinguish between high P-wave, T-wave, QRS-complex and baseline drifts. Some techniques [3] combine wavelet transform and Hidden Markov-model to get precise segmentation and classification including P, QRS and T, waveform detection, identification of normal waves, and premature ventricular contractions. Some techniques [27] detect morphology changes in waveforms using HMM that can classify beats between normal and premature electrical activity based on QT interval analysis [10].

Coast [8] presented arrhythmia analysis using HMM for the first time. The importance of P-wave detection for supraventricular arrhythmia along with QRS detection for ventricular arrhythmia was realized and addressed by using P-wave detection technique [7, 15]. However, these techniques do not precisely diagnose and identify all arrhythmia subclasses due to the insufficient P-wave modeling in HMM. Abnormality of P-wave is one of the important attributes to be considered while classifying supraventricular arrhythmia. None of the previous works identify exhaustively different subclasses of arrhythmia.

P-wave analysis by Clavier [7] combines Markov-model, slope observation and coefficients of ECG wavelet transform. Then P-wave delineation is done to get the positive and negative slopes of the ECG. These are represented as  $P_1$  for the positive slope of atrial activation and  $P_2$  for the negative slope of atrial activation. Then classification is done based on four parameters: time, space, spectral and wavelet entropy. Similarly QRS-complex is broken into  $Q_1$  for positive slope of ventricular activation and  $Q_2$  for negative slope of ventricular depolarization. T-wave is broken into  $T_1$  and  $T_2$  for positive and negative slope of ventricular repolarization. Then Markov-model is developed on state transition probabilities obtained from a database. Their work is limited as they do not take into account P-waves from left and right atria and P-wave inversions.

In our proposed Markov-model, the finer decomposition of P-waves into four different states based upon P-wave splitting and atria traversal captures different possibilities of electric impulse travel in upper chambers. We present a clear graph matching algorithm-based most-probable-path in patient's Markov-model to identify a subclasses of arrhythmia. We extend the work by Clavier [7] by introducing four different states using both left and right atria as well as slopes. We also analyze inverted P-waves in our Markov-models.

## 8 Conclusions and future work

In this paper, we have proposed a new Markov-model to diagnose different subclasses of supraventricular arrhythmia. The trained Markov-models have high sensitivity to identify patient's disease state during real-time monitoring.. The approach when augmented with morphological analysis [10, 27], improved statistical analysis of multiple PTGs from a patient and bigger training database will facilitate automated identification of various subclasses of supraventricular arrhythmia with a high confidence factor. Currently, we are refining Markov-model to include morphological data such as amplitude, interval, inversion of T-wave, and change in

isoelectric line. We are also extending the model to identify subclasses of ventricular arrhythmia [21].

## 9 References

- [1] J. S. Andersen, H. Egeblad, U. Abildgaard, J. Aldershvile and J. Godtfredsen. "Atrial fibrillation and left atrium enlargement: cause or effect?"; *J Intern Med.*, Vol. 229, Issue 3, 253-256, 1991.
- [2] R. V. Andreão, B. Dorizzi and J. Boudy. "ECG signal analysis through hidden Markov-models"; *IEEE Transactions on Biomedical Engineering*, 53(8), 1541-1549, 2006.
- [3] R. V. Andreão and J. Boudy. "Combining Wavelet Transform and Hidden Markov-models for ECG Segmentation"; *EURASIP Journal on Advances in Signal Processing*, Article ID 56215, 95-95, 2007.
- [4] Centers of disease and control prevention. <http://www.cdc.gov/heartdisease/facts.htm>.
- [5] B. Castro, D. Kogan and A. B. Geva. "ECG feature extraction using optimal mother wavelet"; *Proc. of the 21<sup>st</sup> IEEE Convention of Electrical and Electronic Engineers, Israel*, 346-350, 2000.
- [6] W. T. Cheng, and K. L. Chan. "Classification of electrocardiogram using hidden Markov-models"; *Engineering in Medicine and Biology Soc., Proc. of the 20th Annual International Conf. of the IEEE*; 1:143-146, 1998.
- [7] L. Clavier, J. M. Boucher, R. Lepage, J. J. Blanc, and J. C. Cornily. "Automatic P-wave analysis of patients prone to atrial fibrillation"; *Med. Biol. Eng. Comput.*, Volume 40, 63-71, 2002.
- [8] D. A. Coast. "An approach to cardiac arrhythmia analysis using hidden Markov-models"; *IEEE Trans. on Biomedical Engineering*, 37(9):826-836, 1990.
- [9] R. A. Colucci, M. J. Silver, J. Shubrook. "Common types of supraventricular tachycardia: diagnosis and management" *Am Fam Physician*, 82(8), 942-952, 2010.
- [10] C. P. Day, J. M. McComb, R. W. F. Campbell. "QT Dispersion: An Indication of Arrhythmia Risk in Patients with Long QT Intervals"; *British Heart Journal*, (63)6, 342-344, 1990.
- [11] T. B. Garcia and N. E. Holtz. "12-Lead ECG: The Art of Interpretation". Jones and Bartlett, 2013.
- [12] T. B. Garcia and G. T. Miller. "Arrhythmia Recognition: The Art of Interpretation". Jones and Bartlett, 2013.
- [13] C. A. Garcia, A. Otero, X. Vila and D. G. Marquez. "A New Algorithm for Wavelet-based Heart Rate Variability Analysis"; *Biomedical Signal Processing and Control*, (8) 6, 542-550, 2013.
- [14] W. M. Jackman et al. "Treatment of supraventricular tachycardia due to atrioventricular nodal reentry by radiofrequency catheter ablation of slow-pathway conduction"; *New England Journal of Medicine*, 327.5, 313-318, 1992.
- [15] M. J. Janse. "Why does atrial fibrillation occur?"; *European Heart Journal*, 18(Supp. C):C12-C18, 1997.
- [16] M. E. Josephson. "Clinical cardiac electrophysiology: techniques and interpretations". Lippincott Williams & Wilkins, 2008.
- [17] J. Lee, D. D. McManus, P. Bourrell and L. K. H. Sörnmo Chon. "Atrial Flutter and Atrial Tachycardia Detection using Bayesian approach with high resolution time-Frequency spectrum From ECG Recordings"; *Biomedical Signal Processing and Control*, 8(6), 992-999, 2013.
- [18] S. J. Mahmoodabadi, A. Ahmadian and M. D. Abolhasani. "ECG Feature Extraction using Daubechies Wavelets"; *Proc. of fifth IASTED International conference on Visualization, Imaging and Image Processing*, 343-348, 2005.
- [19] J. P. Nolan, R. W. Neumar and C. Adrie. "Post-cardiac Arrest Syndrome: Epidemiology, Pathophysiology, Treatment, and Prognostication"; *Resuscitation*, Vol. 79, 350-379, 2008.
- [20] J. E. Olgin and D. P. Zipes, "Heart Disease: A Textbook of Cardiovascular Medicine". 8th edition, Saunders Elsevier, 2007.
- [21] L. F. Pava, P. Perafán, M. Badiel, J. J. Arango, L. Mont, C. A. Morillo and J. Brugada "R-wave Peak Time at DII: A New Criterion for Differentiating between Wide Complex QRS Tachycardias"; *Heart Rhythm*, 7(7), 922-26, 2010.
- [22] G. Peretto, A. Durante, L. R. Limite, and D. Cianflone, " Postoperative Arrhythmias after Cardiac Surgery: Incidence, Risk Factors, and Therapeutic Management"; *Cardiology Research and Practice*, Volume 2014, Article ID 615987, 15 pages, 2014.
- [23] The Physionet Database for ECG. <http://www.physionet.org/physiobank/database/#ecg> <http://www.physionet.org/physiobank/database/svdb/>
- [24] S. Russell and P. Norvig. "Artificial Intelligence". third edition, Prentice-Hall, 2010.
- [25] St. Petersburg's Institute of cardiological technics 12 lead arrhythmia database, available at <http://physionet.org/physiobank/database/incartdb/>
- [26] C. D. A. and R. M. Allegheny. An approach to cardiac arrhythmia analysis through hidden Markov-model"; *IEEE Trans. of Biomedical Engineering*, 37(9), 826-836, 1990.
- [27] P. Tadejko and W. Rakowski. "Mathematical Morphology Based ECG Feature Extraction for the Purpose of Heartbeat Classification"; *Proc. of the 6<sup>th</sup> International Conference on Computer Information Systems and Industrial Management Applications*, 322-327, 2007.
- [28] The WFDB software package, available at <http://www.physionet.org/physiotools/wfdb.shtml#downloading>



# A Belief Rule Based Expert System to Diagnose Measles under Uncertainty

Md. S. Hossain<sup>1</sup>, K. Andersson<sup>2</sup>, and S. Naznin<sup>1</sup>

<sup>1</sup>Computer Science and Engineering, University of Chittagong, Chittagong-4331, Bangladesh

<sup>2</sup>Pervasive and Mobile Computing Laboratory, Luleå University of Technology, S-931 87 Skellefteå, Sweden

**Abstract** - Measles is a highly infectious child disease that causes serious complications and death worldwide. Measles is generally diagnosed from its signs and symptoms by a physician, which cannot be measured with 100% certainty during the diagnosis process. Consequently, the traditional way of diagnosing measles from its signs and symptoms lacks the accuracy. Therefore, a belief rule-based inference methodology using evidential reasoning approach (RIMER), which is capable of handling various types of uncertainties has been used to develop an expert system to diagnose measles under uncertainty. The results, generated, from the system have been compared with the expert opinion as well as with a Fuzzy Logic based system. In both the cases, it has been found that the Belief Rule Based Expert (BRBES), presented in this paper, is more reliable and accurate.

**Keywords:** Measles, Expert System, Belief Rule Base, Uncertainty, Evidential Reasoning

## 1 Introduction

Measles is a highly contagious infection caused by measles virus. It is a common disease in developing countries like Bangladesh. Almost every organ system can be affected by the complications from measles [1]. Physical contact, coughing and sneezing can spread the infection. Moreover, infected droplets of mucus can remain active and contagious for around two hours. This means that the virus can live outside the body - for example, on surfaces, pens, pencils etc. A survey said that, before the introduction of measles vaccines, the disease occurred in 95%–98% of children by the age of 18 years [2][3]. After an incubation period of 8–12 days, measles begins with increasing fever (to 39C–40.5C) and cough, coryza, and conjunctivitis [4][5]. Symptoms intensify over the 2–4 days before the onset of rash and peak on the first day of rash [6]. The rash is usually first noted on the face and neck, appearing as discrete erythematous patches 3–8 mm in diameter. The lesions increase in number for 2 or 3 days, especially on the trunk and the face, where they frequently become confluent. Discrete lesions are usually seen on the distal extremities, and with careful observation, small numbers of lesions can be found on the palms of 25%–50% of those infected. The rash lasts for 3–7 days and then fades in the same manner as it appeared, sometimes ending with a fine desquamation that may go unnoticed in children who are bathed daily. An exaggerated desquamation is commonly seen in malnourished children [5][7][8]. Fever usually persists for 2 or 3 days after the onset of the rash, and

the cough may persist for as many as 10 days. Koplik's spots usually appear 1 day before the onset of rash and persist for 2 or 3 days. These bluish-white, slightly raised, 2- to 3-mm-diameter lesions on an erythematous base appear on the buccal mucosa, usually opposite the first molar, and occasionally on the soft palate, conjunctiva, and vaginal mucosa [9][10]. Koplik's spots have been reported in 60%–70% of persons with measles but are probably present in most persons who develop measles [11]. An irregular blotchy enanthem may be present in other areas of the buccal mucosa. Photophobia from iridocyclitis, sore throat, headache, abdominal pain, and generalized mild lymphadenopathy are also common. Measles is transmitted by the respiratory route and is highly infectious. Infectivity is greatest in the 3 days before the onset of rash, and 75%–90% of susceptible household contacts develop the disease [12][13][14]. The early pre-rash symptoms are similar to those of other common respiratory illnesses, and affected persons often participate in routine social activities, facilitating transmission. The outbreaks of measles occur when children in the first few days of illness attend school and other events like sports [15]. Outbreaks also occur when ill children are brought to a doctor's office or emergency room for evaluation for fever, irritability, or rash [16][17].

Medical diagnosis is a process to detect a disease by measuring its specific signs and symptoms. Signs are observed by a physician while symptoms are expressed by the patient. However, patients may not be able to precisely express their feelings on the symptoms and doctors may not always be sure about the condition of signs. Therefore, various types of uncertainties may be noticed during the diagnosis process of a disease. These uncertainties should be carefully handled, enabling the accurate diagnosis of a disease.

The main signs and symptoms associated with measles consist of fever, cough, coryza, conjunctivitis, koplik's spot and rash [18] [19]. These signs and symptoms cannot be measured with 100% certainty during the diagnosis process. The reason for this is that they are expressed both by the patients and the doctors in linguistic terms such as "high", "medium" and "low". Such linguistic terms inevitably contain uncertainty such as vagueness, imprecision and ambiguity. Sometime, it is not possible to obtain data on the above signs and symptoms and hence, this causes uncertainty due to ignorance and incompleteness in the diagnosis process. Consequently, the traditional way of diagnosing the measles

is incapable of producing the accurate diagnostic results because of the absence of appropriate method of handling the issue of uncertainty that exists with the signs and symptoms. Problem of this nature could be well handled by developing an expert system [20], which emulate the human decision making process. However, this system employed traditional IF-THEN rules to develop knowledge base, which can capture assertive knowledge not the uncertain knowledge. Therefore, a knowledge representation schema is required which could be used to develop the expert system with the capability of capturing different types of uncertainties that exist with the signs and symptoms of measles. Hence, belief rule base knowledge representation schema, which has the ability to represent uncertain knowledge, has been considered to develop an expert system along with the evidential reasoning as the inference engine [21].

The remaining of the paper is structured as follows. Section two presents the related works. Section three provides an overview of the methodology used to develop the belief rule based expert system (BRBES) known as RIMER. Section four describes the design and implementation of the BRBES. Section five presents the results and discussions. Section six concludes the paper.

## 2 Literature Review

Expert systems are widely used to diagnose diseases. This is a computer system that nearly earns the decision making ability of a human expert. The goal of designing an expert system is to solve a complex problem by reasoning with knowledge rather than by conventional procedural codes. An expert system consists of two parts namely, the inference engine and the knowledge base [22]. Knowledge base of an expert system dedicated to represent the knowledge about a domain. Various knowledge representation schemas such as Propositional Logic (PL), First Order Logic (FOL), Fuzzy Logic (FL), Frames, Semantic Net, Case Based Reasoning and Bayesian Belief Network are widely used to build the knowledge base of an expert system. Both PL and FOL are used to acquire assertive knowledge and hence, unable to represent various types of uncertainties [23]. Semantic Net is a directed or undirected graph, which represents semantic relations between concepts [24] but not the uncertainty. Bayesian Belief Network or BNN is a directed acyclic graph representing the facts using conditional probability. Given relationships between diseases and symptoms, the network can compute the probabilities of the presence of various diseases. The idea of conditional probability is very useful here. In spite of remarkable power of BNN it has some limitations such as it is difficult to compute result for previously unknown network. A Bayesian network is only useful when a priori knowledge is reliable [25]. However, the diagnosing of measles requires the consideration of a posteriori signs and symptoms. Fuzzy logic is suitable to represent lexical knowledge and can address uncertainties such as vagueness, imprecision and ambiguity but is not equipped enough to handle uncertainty due to ignorance or incompleteness. Belief rule-base [21] is a new

knowledge representation schema uses a belief structure where belief degrees are embedded with all possible terms of the consequent part of a rule. It can handle various types of uncertainties such as vagueness, ambiguity, imprecision, ignorance and incompleteness.

Inference is the process of reasoning from the factual knowledge in order to derive logical conclusions. Examples of inference procedures are Forward Chaining (FC), Backward Chaining (BC), Modus Ponens (MP), evidential reasoning(ER) etc. MP is the simplest form of inference procedure, enables the deductive reasoning. It is interesting to note that both FC and BC developed based on MP. Forward chaining begins with the available fact or input and searches the rules until the rule is found where the if-clause is known to be true. When such a rule is found, the process may conclude or forward to the then-clause for further inference. The process is iterated until a goal is reached. On the other hand backward chaining works backward from the goal. It begins with a list of goals and searches the rules until it finds a rule, which contains a desired goal in then-clause. BC is a goal driven mechanism while FC is a data driven mechanism for inference. However, they are not equipped with the mechanism to handle uncertainties. On the other hand Evidential Reasoning, which is used as an inference mechanism can process various types of uncertainties mentioned previously, especially the ignorance and incompleteness. Therefore, belief rule base along with evidential reasoning, known as RIMER, has been considered to develop the expert system to diagnose measles under uncertainty.

Therefore, in order to ensure a most correct detection of measles under uncertainty, an expert system needs to be developed by using BRB for knowledge base, along with the ER approach for inference mechanism.

## 3 Overview of RIMER

RIMER consists of two parts. The first part is concerned with building the knowledge base using BRB while the second part is concerned with the inference procedures consisting of input transformation, rule activation weight calculation, belief degree update and rule aggregation using evidential reasoning.

### 3.1 Domain knowledge representation using BRB

BRB is an extended form of traditional IF-Then rule. In a BRB, the antecedent part consists of antecedent attributes that take referential values, while the referential values of the consequent of the consequent part is attached with belief degrees. In this way, the later part forms a belief structure as shown in equation (1). The BRB also includes learning parameter such as rule weight and antecedent attribute weight.

$R_k$ :  $if(S_{k1} \text{ is } A_1^k) \wedge (S_{k2} \text{ is } A_2^k) \wedge \dots \wedge (S_{kT_k} \text{ is } A_{T_k}^k)$ ,  
 then  
 $\{(D_1 \text{ is } \beta_{1k}), (D_2 \text{ is } \beta_{2k}), \dots, (D_N \text{ is } \beta_{Nk})\}$   
 with a rule weight  $\theta_k$  and  
 attribute weights  $\delta_{k1}, \delta_{k2}, \dots, \delta_{kT_k}$   
 $k \in \{1, \dots, L\} \dots (1)$

Here,  $S_{k1}, S_{k2}, \dots, S_{kT_k}$  stand for the antecedent attributes in  $k^{th}$  rule and each  $A_j^k (j \in \{1, \dots, T_k\})$  is one of the referential values of  $S_{ki}$ .  $T_k$  is the total number of antecedent attributes in the  $k^{th}$  rule.  $D_j$  is one of the consequent reference values of the belief rule.  $\beta_{jk} (j = 1, \dots, N, k = 1, \dots, L)$  is the degree of belief to which the consequent reference value  $D_j$  is believed to be true. If  $\sum_{j=1}^N \beta_{jk} = 1$  the  $k$ th rule said to be complete;  $L$  number of all belief rules in the rule base.  $N$  is the number of all possible consequent in the rule base. For example, in the case of measles diagnosis, a belief rule can be written as follows:

$R_k$ :  $if(\text{Fever is 'Medium'}) \wedge (\text{Rash is 'High'}) \wedge$

Here,  $\{(High, 0.90), (Medium, 0.10), (Low, 0.00)\}$  is a belief distribution associated with the measles diagnosis. This belief distribution states that, it is 90% sure that the condition of the measles is 'high' and 10% sure that it is 'medium'. In this belief rule, the total degree of belief is  $(0.90+0.10+0) = 1$  and hence, the assessment is complete.

### 3.2 Inference Procedures

Following the schema of BRB as represented by equation (1) the knowledge base of an expert system is developed, which is known as initial rule base. This rule base is stored in the static memory or in the long-term memory. However, in order to diagnose the measles, it is necessary to get its values of the signs and symptoms, collected either from the patients or from the physicians. These values considered as the input values of the BRBES. Basically, these signs and symptoms considered as the antecedent attributes of a rule. Each input value needs to be distributed over the referential values of an antecedent attribute to demonstrate what amount of this input value match with each of the referential value. This is shown in equation (2). The referential values of each antecedent attribute may by "High", "Medium" and "Low", which are similar to the signs and symptoms of the measles as express by the patients or the physicians in terms of these linguistic terms.

$$(S_{k1} \text{ is } A_1^k, \alpha_1^k) \wedge (S_{k2} \text{ is } A_2^k, \alpha_2^k) \wedge \dots \wedge (S_{kT_k} \text{ is } A_{T_k}^k, \alpha_{T_k}^k) \dots (2)$$

where  $\alpha_i^k$  is the matching degree of the input value to the referential value  $A_i^k$  of an antecedent attribute  $S_{ki}$  in the  $k^{th}$  rule ( $i = 1, \dots, T_k$ ). This is important to note that when the matching degree is assigned to the referential values of the attributes of a rule then it is said to be activated. This phenomenon is also called packet antecedent of a rule. In

other way, it can be argued that the rule is in the RAM or on the short term memory. The total degree or the combined matching degree  $\alpha_k$ , to which the input matches the whole antecedent part of  $k^{th}$  rule, can be calculated by using the following formula:

$$\alpha_k = aggr((\delta_{k1}, \alpha_1^k), \dots, (\delta_{kT_k}, \alpha_{T_k}^k)) \dots (3)$$

where  $aggr$  is an aggregation function which should be selected carefully. In the case of a rule defined as in (3), following simple weighted multiplicative aggregation function can be used:

$$\alpha_k = \prod_{i=1}^{T_k} (\alpha_i^k)^{\bar{\delta}_{ki}} \dots (4)$$

where  $\bar{\delta}_{ki} = \frac{\delta_{ki}}{\max_{i=1, \dots, T_k} \{\delta_{ki}\}}$  so that  $0 \leq \bar{\delta}_{ki} \leq 1$ .

The activation weight  $w_k$  for  $k^{th}$  rule can be generated by the following equation:

$$w_k = \theta_k \alpha_k / \sum_{i=1}^L \theta_i \alpha_i \dots (5)$$

This activation weight will be zero if the  $k^{th}$  rule is not activated.

The incompleteness of the consequent of a rule can also be occurred by its antecedents because of the lack of data. Such incompleteness should be taken into account in the inference process. Therefore, the existing belief degree is updated using the actual input information.

$$\beta_{ik} = \bar{\beta}_{ik} \sum_{t=1}^{T_k} (\lambda(t, k) \sum_{j=1}^{I_t} \alpha_{tj}) / \sum_{t=1}^{T_k} \lambda(t, k) \dots (6)$$

where

$$\lambda(t, k) = \begin{cases} 1, & \text{if } t^{th} \text{ attribute is used in defining } R_k \\ 0, & \text{otherwise.} \end{cases}$$

ER approach is used to aggregate all the packet antecedents of the  $L$  rules to obtain the degree of belief of each referential values of the consequent attribute by taking account of given input values  $P_i$  of antecedent attributes. This aggregation can be carried out either using recursive or analytical approach. Accordingly, for each  $D_i$  the generated belief degree will be calculated as:

$$\beta_i = \frac{P_{iJ(L)}}{1 - \bar{P}_{D_i(L)}}, i = 1, \dots, N \dots (7)$$

Thus, the final result generated by aggregating all rules can be represented as

$$F = \{D_i, \beta_i\}; i = 1, \dots, N \dots (8)$$

The entire result can then be converted into a crisp value by calculating the expected score  $u(F)$ .

$$u(F) = \sum_{i=1}^N \beta_i u(D_i) \quad \dots (9)$$

$u(D_i)$  denotes the score of an individual consequent  $D_i$ .

### 4 BRBES for Measles Diagnosis

This section presents the design, implementation, knowledge-base construction, interface of the BRBES to diagnose measles.

#### 4.1. Architecture, design and Implementation of the BRBES

The design of a system consists of data structure and program components that are essential to build a computer based system. It also considers the system organization pattern, which is known as architectural style. The architecture of the BRBES consists of user interface, data management layer and application processing as shown in Figure 1. User interface interacts to system user to get input data and to receive system generated output. Data management layer includes designing the knowledge base containing the rules and facts about measles.

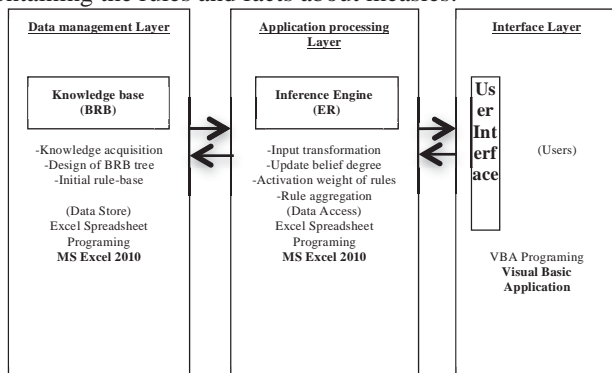


Figure 1. Architecture of BRBES

The application processing layer is concerned with input transformation, rule activation weight calculation, belief degree update and the aggregation of the rules. The data management layer and application processing layer has been implemented using Microsoft Excel 2010 spreadsheet programming. The interface layer has been implemented using VBA (visual basic application) programming [27][28][29][30][31][32].

#### 4.2. Knowledgebase Construction in BRB

Figure 2 illustrates the BRB tree, which has been developed in consultation with the physicians. The BRB tree represents a multilevel hierarchical structure of the knowledge base. Each sub tree with a root and its leaf nodes represents a sub-rule-base of the knowledge base.

The leaf nodes becomes the antecedent attributes of a rule and the parent node becomes its consequent attribute. The number of sub rule-bases can be determined from the number of parent nodes. The BRB tree as illustrated in Figure2 has two parent nodes i.e.  $A_8$  and  $A_4$ . Therefore, there should be two sub-rule-bases namely sub-rule base  $A_4$  and sub-rule-base  $A_8$ .

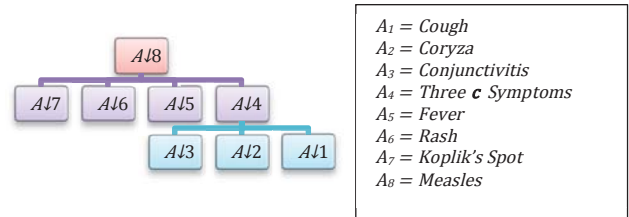


Figure 2. The BRB Framework to Diagnose Measles

#### Sub-Rule-Base $A_4$

This sub rule base consist of three antecedent attributes namely  $A_1$ ,  $A_2$  and  $A_3$ . Each attribute has three referential values, namely high (H), medium (M) and low (L). Therefore, this sub-rule-base consists of 27 rules.

#### Sub-Rule Base $A_8$

Four antecedent attributes  $A_4$ ,  $A_5$ ,  $A_6$  and  $A_7$  exist in this sub-rule-base. Three of them consist of three referential values high (H), medium (M) and low (L). However,  $A_7$  has two referential values such as high (H) and low (L). Therefore this sub-rule-base should have  $27 \times 2 = 54$  rules. Table 1 illustrates the initial rule base for the sub rule base  $A_8$ .

TABLE 1. INITIAL SUB-RULE-BASE  $A_8$

Rul e id	Rule Weigh t	IF	THEN
1	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is H ^ $A_4$ is H	$A_8$ is {(H,1.00), (M,0.00), (L,0.00)}
2	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is H ^ $A_4$ is M	$A_8$ is {(H,0.9), (M,0.10), (L,0.00)}
3	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is H ^ $A_4$ is L	$A_8$ is {(H,0.80), (M,0.20), (L,0.00)}
4	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is M ^ $A_4$ is H	$A_8$ is {(H,0.9), (M,0.1), (L,0)}
5	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is M ^ $A_4$ is M	$A_8$ is {(H,0.5), (M,0.5), (L,0.00)}
6	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is M ^ $A_4$ is L	$A_8$ is {(H,0.2), (M,0.8), (L,0.00)}
7	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is L ^ $A_4$ is H	$A_8$ is {(H,0.8), (M,0.2), (L,0)}
8	1	$A_7$ is H ^ $A_6$ is H ^ $A_5$ is L ^ $A_4$ is M	$A_8$ is {(H,0.2), (M,0.8), (L,0)}
...	...	...	...
54	1	$A_7$ is L ^ $A_6$ is L ^ $A_5$ is L ^ $A_4$ is L	$A_8$ is {(H,0.00), (M,0.00), (L,1.00)}

An example of a belief rule taken from Table 1 is illustrated below.

*R1: IF Koplik's Spot is High AND Rash is high AND Fever is High AND 'Three C Symptoms' is High THEN Measles{(High,1),(Medium,0),(Low,0)}.*

In the above belief rule, the belief degrees are attached to the three referential values. The weight of each rule has been considered as '1'.

### 4.3. BRBES Interface

A system interface can be defined as the media, enabling the interaction between the users and the system. Figure 3 illustrates a simple interface of the BRBES. This interface facilitates the acquiring of the leaf nodes (antecedent attributes) data of the BRB framework (fig 2). The system interface enables the displaying of the measles diagnosing results (the top node). For example, Figure 3 illustrates the result for the data of leaf nodes (A1 = mild, A2 = moderate, A3 = mild, which appears in the signs and symptoms dialogue box of the BRBES' interface) associated with sub-rule-base A4.

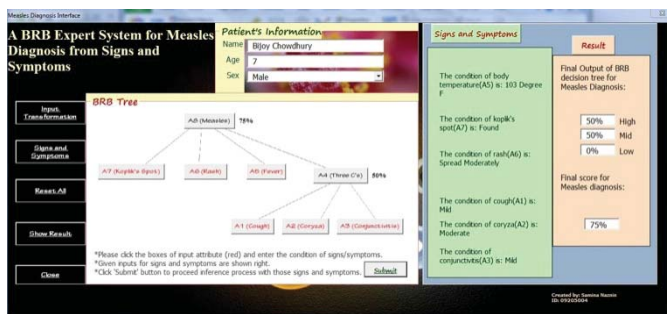


Figure 3. User Interface of BRBES to Diagnose Measles

From Figure 3 it can be observed that the assessment of 'Three C Symptoms', which is the consequent attribute named as "A4" of this sub-rule-base is 50%. This data has been calculated by the system by applying equations (7-9). The child node of the sub rule base A8 consists of A4, A5, A6, and A7. However, A5, A6 and A7 are the leaf nodes and the input data related to these nodes as shown in the signs and symptoms dialogue box of the system interface are '103 degree F', 'Spread Moderately' and 'Found' respectively. These linguistic data have been distributed over the referential values of the leaf node attributes by using equations (2). Figure 3 illustrates the overall assessment of the measles which is 50% high, 50% medium and 0% low. This is obtained by using equation (8) while by using equation (9) the assessment in terms of crisp value i.e. 75% has been achieved.

## 5 Result and Discussion

In this research, the leaf nodes data of the BRB framework (fig 2) have been collected from the patients. The patient's data have been used in the BRBES to diagnose measles. Expert's opinion on the measles diagnosis of the patients is also collected as shown in Table 3. The data set consists of 200. For simplicity, only ten patient's data set are presented in Table 2. Colum 8 of Table 2 illustrates BRBES" generated output in percentages, which is calculated by using utility equation (9). For example, the overall system output of measles diagnosis is 75.28% can be obtained, by using a

degree of belief associated with referential values such as {High (0.6191), Medium(0.2674), Low(0.1134)}.

The Receiver Operating Characteristic (ROC) curve can be used to analyze effectively performances of diagnosis of measles having ordinal or continuous results [33][34][35][36][37]. It can be used to test the reliability of the BRBES's output in comparison with manual system by taking account of benchmark data. The real laboratory test results of the patients have been considered as the benchmark data. For example, if the measles is found positive then it is considered as "1", otherwise it is "0" as shown in the column 10 of the Table 2. The performance of the system can be measured by calculating the Area under Curve (AUC). A larger AUC refers to a more accurate and reliable result. Fig 4 shows the two ROC curves; one represents the performances of the BRBES and the other for the manual system. The performance of the BRBES has also been compared against fuzzy rule based expert system, as shown in Fig 5.

TABLE II. DATASET FOR SYSTEM TESTING

S.n o.	A1	A2	A3	A5	A6	A7	Syste m Result (%)	Expert' s opinion (%)	State varia ble
1	Mo derate	Mild	Mild	103.0	Spread moderately	Found	82.00	80.00	1
2	Mi ld	Normal	Norm al	102.0	Spread moderately	Found	67.50	65.00	1
3	No rm al	Normal	Mode rate	99.5	Spread moderately	Found	64.50	60.00	1
4	Mi ld	Mild	Norm al	102.5	Spread moderately	Found	72.50	70.00	1
5	No rm al	Normal	Norm al	99.5	A few rashes	Not found	1.50	1.00	0
6	Mi ld	Normal	Norm al	102.5	A few rashes	Found	38.00	35.00	1
7	No rm al	Severe	Norm al	103.5	A few rashes	Found	65.50	65.00	1
8	Mi ld	Normal	Norm al	102.0	A few rashes	Found	38.00	35.00	1
9	No rm al	Severe	Norm al	104.0	A few rashes	Found	65.50	60.00	1
10	No rm al	Normal	Norm al	99.5	A few rashes	Not found	1.50	1.50	0
11	Mi ld	Normal	Norm al	98.4	A few rashes	Not found	4.00	4.00	0
12	No rm al	Normal	Norm al	101.5	A few rashes	Found	27.50	27.00	0
13	No rm al	Normal	Mode rate	99.5	Spread moderately	Found	64.50	60.00	1
14	Mi ld	Mild	Norm al	102.5	Spread moderately	Found	72.50	65.00	1
15	Mi ld	Normal	Norm al	101.5	Spread moderately	Found	67.50	60.00	1

### 5.1. Reliability of BRBES against Expert Opinion

The ROC curve plotted by the blue line in fig 6 is carried out by using the results generated by the BRBES, which AUC is 0.849 (95% confidence intervals 0.729 - 0.970). The ROC curve plotted by the green line in fig 6 is obtained from the result set manually produced based on the physician's opinion, and its AUC is 0.811 (95% confidence intervals 0.675 - 0.947). The 95% confidence interval means that the parameter value in estimation can remain within the range of estimated interval with 95% certainty [38]. The results show that the AUC of BRBES is greater than that of physician

opinion. This implies that, under clinical uncertainties, the diagnostic performance of the BRBES is better than manual judgment made by a physician.

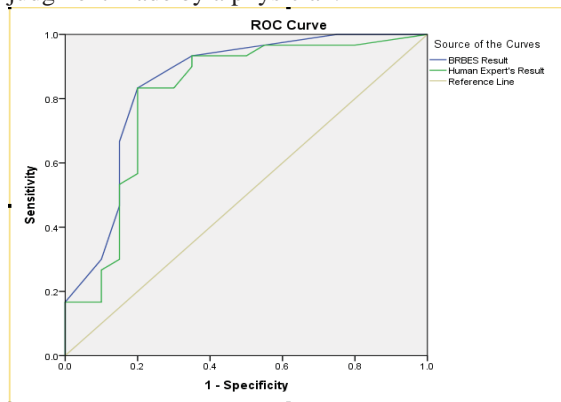


Figure 4. ROC curves of Measles Diagnosis between BRBES and Expert Opinion

## 5.2. Comparison of BRBES with Fuzzy Logic based System

The BRBES has been compared with a fuzzy rule based expert system. The main difference between fuzzy rule and the belief rule can be understood from the representation of R1 by using these two knowledge representation schemas.

*R1: IF Koplik's Spot is High AND Rash is high AND Fever is High AND 'Three C Symptoms' is High THEN Measles(High,1),(Medium,0),(Low,0)}. [Belief Rule]*

*R1: IF Koplik's Spot is High AND Rash is high AND Fever is High AND 'Three C Symptoms' is High THEN Measles is High [Fuzzy Rule].*

From the above, it can be seen that in case of fuzzy logic degree of belief is not embedded with the consequent part of the rule. The same patients data (which were used with both BRBES and expert) are used to obtain the results of measles diagnosis by using fuzzy logic based expert system (FLBES). Table 3 illustrates the comparison of measles diagnosis results among BRBES, FLBES and expert opinion.

TABLE III COMPARISON AMONG BRBES, FLBES AND EXPERT OPINION

S.N o.	BRBES (%)	Expert opinion (%)	FLBES (%)	Benchmark Data
1	82.00	80.00	80.09	1
2	1.50	1.50	10.02	0
3	64.50	60.00	60.00	1
4	72.50	70.00	80.09	1
5	1.50	1.00	10.02	0
6	38.00	35.00	29.80	1
7	65.50	65.00	60.00	1
8	38.00	35.00	29.80	1
9	65.50	60.00	60.00	1
10	67.50	65.00	80.09	1
11	4.00	4.00	10.02	0
12	27.50	27.00	29.80	0
13	64.50	60.00	60.00	1
14	72.50	65.00	80.09	1
15	67.50	60.00	80.09	1

The ROC curve plotted by red line in fig 5 is obtained by using FLBES, which AUC is 0.824 (95% confidence intervals 0.693 - 0.956). The ROC curve plotted by the blue line is obtained by using result generated by the BRBES, and its AUC is 0.849 (95% confidence intervals 0.729 - 0.970). The ROC curve plotted by the green line is obtained from the result set manually produced based on the expert opinion, which AUC is 0.811 (95% confidence intervals 0.675 - 0.947). The results demonstrate that the AUC of BRBES is greater than that of fuzzy rule based expert system as well as that of expert opinion. This implies that, under clinical uncertainties, the diagnostic performance of the BRBES is better than that of FLBES.

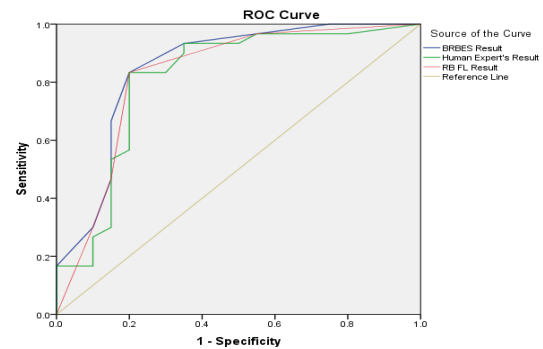


Figure 5. ROC curves of Measles Diagnosis among BRBES, Fuzzy Logic Based Expert System and Expert Opinion

## 6 Conclusion

The development and application of a BRBES to diagnose measles under uncertainty presented in this paper. The system employed RIMER methodology, which is capable of handling of various types of uncertainties found in the clinical domain knowledge as well as in the signs and symptoms. The results of the BRBES compared against expert opinion and found more reliable and robust. The reason for this is that expert is unable to consider various types of uncertainties such as vagueness, imprecision, ambiguity, randomness and ignorance, those are associated with the signs and symptoms of the measles during the diagnostic process. In addition, the performance of the BRBES has been compared against fuzzy logic based expert system. It is observed that the results generated from BRBES are better than fuzzy based expert system (fig 5). The reason for this is that fuzzy logic only considers uncertainties due to vagueness, imprecision and ambiguity while BRBES in addition to these uncertainties considers uncertainty due to randomness and ignorance. In addition, the inference procedures of BRBES consists of input transformation, rule activation weight calculation, belief update and rule aggregation using evidential reasoning approach. Evidential reasoning is capable of process various types of uncertainties, which is not the case with the fuzzy based inference engine such as Mandni and Takagi-Sugeno (TS).

## 7 References

- [1] T. P. Robert and A. H. Neal, "The Clinical Significance of Measles: A Review," *The Journal of Infectious Diseases*, vol. 189, pp. 4-16, 2004.
- [2] A. Langmuir, "Medical importance of measles," *Am J Dis Child*, vol. 103, pp. 224-6, Mar. 1962.
- [3] M. Snyder, F. McCrumb, T. Bigbee, A. Schluederberg, and Y. Togo, "Observations on the seroepidemiology of measles," *Am J Dis Child*, vol. 103, p. 250-1, 1962.
- [4] S. Krugman, et al., Eds., "Measles (rubeola)," in *Infectious disease of children*. Saint Louis: Mosby Year Book, 1992, p. 223-245.
- [5] F. Robbins, "Measles: clinical features," *Am J Dis Child*, vol. 103, pp. 266-273, 1962.
- [6] S. Krugman, "Further-attenuated measles vaccine: characteristics and use," *Rev Infect Dis*, vol. 5, p. 477-81, 1983.
- [7] D. Scheifele and C. Forbes, "Prolonged giant cell excretion in severe African measles," *Pediatrics*, vol. 50, pp. 867-873, 1972.
- [8] D. Morley, "Measles in the developing world," *Proc R Soc Med*, vol. 67, pp. 1112-5, 1974.
- [9] H. Koplik, "The diagnosis of the invasion of measles from a study of the exanthema as it appears on the buccal mucosa," *Arch Pediatr*, vol. 13, pp. 918-22, 1896.
- [10] D. Suringa, L. Bank, and A. Ackerman, "Role of measles virus in skin lesions and Koplik's spots," *N Engl J Med*, vol. 283, pp. 1139-42, 1970.
- [11] F. Babbott and J. Gordon, "Modern measles," *Am J Med Sci*, vol. 228, pp. 334-61, 1954.
- [12] C. Chapin, "Measles in Providence, Rhode Island, 1858-1923," *Am J Hyg*, vol. 5, pp. 635-55, 1925.
- [13] F. Top, "Measles in Detroit, 1935 -I, Factors Influencing the Secondary Attack Rate Among susceptibles at risk," *Am J Public Health*, vol. 28, pp. 935-43, 1938.
- [14] S. R. Hope, "Infectiousness of communicable diseases in the household (measles, chicken pox, and mumps)," *Lancet*, vol. 2, pp. 549-54, 1952.
- [15] Centers for Disease Control and Prevention, "Measles outbreak among school-aged children," *MMWR Morb Mortal Wkly Rep*, vol. 45, pp. 777-80, 1996.
- [16] A. Bloch, W. Orenstein, and W. Ewing, "Measles outbreak in a pediatric practice: airborne transmission in an office setting," *Pediatrics*, vol. 75, pp. 676-83, 1985.
- [17] P. Remington, W. Hall, I. Davis, A. Herald, and R. Gunn, "Airborne transmission of measles in a physician's office," *JAMA*, vol. 253, pp. 1574-7, 1985.
- [18] D. Baxby, "The diagnosis of the invasion of measles from a study of the exanthema as it appears on the buccal mucous membrane," *Reviews in Medical Virology*, vol. 7, pp. 71-74, Jul. 1997.
- [19] "Bug of the Month - Measles," *Banner Gateway Medical Center*, May 2012.
- [20] F. Basciftci and M. Hacimurtazaoglu, "Simplified Rules Base Obtained with Logic Minimization Method for Diagnosis of Measles Disease Realized with Expert Systems," in *ICSOF*, 2009.
- [21] J. B. Yang, J. Liu, J. Wang, H.-S. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-RIMER," *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans*, vol. 36, pp. 266-285, 2006.
- [22] S. Rahaman, Md. M. Islam, and Md. S. Hossain, "A Belief Rule Based Clinical Decision Support System Framework," in *17th Int'l Conf. on Computer and Information Technology*, Dhaka, 2014, pp. 165-169.
- [23] H. Levesque and R. Brachman, "A Fundamental Tradeoff in Knowledge Representation and Reasoning," *Reading in Knowledge Representation*, p. 41-70, 1985.
- [24] J. F. Sowa, "Semantic Networks," in *Encyclopedia of Artificial Intelligence*, 1987.
- [25] M. L. Vaněk, *Introduction into Bayesian Network*.
- [26] J. Bih, "Paradigm Shift: An Introduction to Fuzzy Logic," *IEEE Potentials*, pp. 10-21, 2006.
- [27] R. De Levie, *Advanced Excel for scientific data analysis*. Oxford University Press, 2004.
- [28] D. M. Bourg, *Excel scientific and engineering cookbook*. O'Reilly, 2006.
- [29] M. M. H. Seref and R. K. Ahuja, "A portfolio management and optimization spreadsheet DSS," in *Handbook on Decision Support Systems 1: Basic Themes*. Springer, 2008.
- [30] E. Wells and S. Harshbarger, *Microsoft Excel 97 Developer's Handbook*. Microsoft Press, 1997.
- [31] D. L. Harnett and J. F. Horrell, *Data, statistics, and decision models with Excel*. Wiley, 1998.
- [32] H. Austerlitz, *Data acquisition techniques using PCs*, 2nd ed. Academic Press, 2003.
- [33] C. E. Metz, "Basic principles of ROC analysis," *Seminars. Nuclear Med*, vol. 8, p. 283-298, 1978.
- [34] R. Body, "Clinical decision rules to enable exclusion of acute coronary syndromes in the emergency department," Thesis, 2009.
- [35] M. S. Hossain, M. S. Khalid, S. Akter, and S. Dey, "A belief rule based expert system to diagnosed Influenza," *9th International Forum on Strategic Technology (IFOST)*, pp. 113-116, 2014.
- [36] S. Rahman and M. S. Hossain, "A belief rule based clinical decision support system to assess suspicion of heart failure from signs, symptoms and risk factors," *International Conference on Informatics, Electronics & Vision (ICIEV)*, pp. 1-6, 2013.
- [37] G. Kong, "An Online Belief Rule-based Group Clinical Decision Support System," A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy, 2011.
- [38] A. D. Aczel and J. Sounderpandian, *Business Statistics*. London: McGraw Hill, 2005.

# Enhancing Informative Frame Filtering by Water and Bubble Detection in Colonoscopy Videos

Ashok Dahal<sup>1</sup>, JungHwan Oh<sup>1</sup>, Wallapak Tavanapong<sup>2</sup>, Johnny Wong<sup>2</sup>, and Piet C. de Groen<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, U.S.A.

<sup>2</sup>Computer Science Department, Iowa State University, Ames, IA 50011, U.S.A.

<sup>3</sup>Mayo Clinic College of Medicine, Rochester, MN 55905, U.S.A.

**Abstract** - Colonoscopy has contributed to a marked decline in the number of colorectal cancer related deaths. However, recent data suggest that there is a significant (4-12%) miss-rate for the detection of even large polyps and cancers. To address this, we have been investigating an 'automated feedback system' which informs the endoscopist of possible sub-optimal inspection during colonoscopy. A fundamental step of this system is to distinguish non-informative frames from informative ones. Existing methods for this cannot classify water/bubble frames as non-informative even though they do not carry any useful visual information of the colon mucosa. In this paper, we propose a novel texture feature based on accumulation of pixel differences, which can detect water and bubble frames with very high accuracy with significantly less processing time. The experimental results show the proposed feature can achieve more than 93% overall accuracy in almost half of the processing time the existing methods take.

**Keywords:** Colonoscopy; Clustering; Texture; Pixel Difference; Feature Extraction

## 1 Introduction

Colonoscopy is an endoscopic technique that allows a physician to inspect the mucosa of the human colon. It has contributed to a marked decline in the number of colorectal cancer related deaths [1]. However, recent data suggest that there is a significant (4-12%) miss-rate for the detection of even large polyps and cancers [2]. To address this, we have been investigating an 'automated feedback system' which informs the endoscopist of possible sub-optimal inspection during colonoscopy in order to improve the quality of the actual procedure being performed [3, 4].

A fundamental step of this system is to distinguish non-informative frames from informative ones. An informative frame in a colonoscopy video can be broadly defined as a frame which is useful for convenient naked-eye analysis of the colon mucosa (Fig. 1). A non-informative frame has the opposite definition (Fig. 2). In general, non-informative frames can be considered out-of-focus frames. Informative and non-informative frames can be loosely termed as clear and blurry frames, respectively. We developed an accurate algorithm for this informative frame filtering (IFF) [5], which is firstly to detect the presence of such vivid lines, and secondly to measure the amount of curvaceous connectivity they possess.

Then, with a carefully chosen threshold, we identify frames which exhibit more curvaceous connectivity and classify them as informative, and vice-versa.

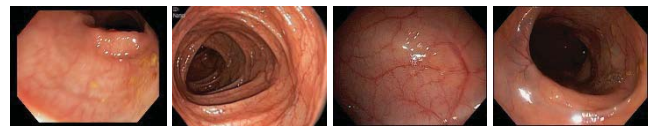


Fig. 1. Examples of Informative Frames.

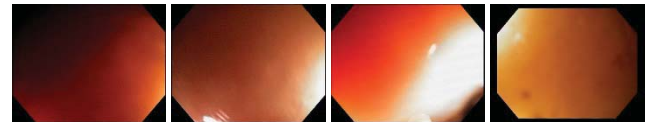


Fig. 2. Examples of Non-Informative Frames.

Fig. 3 shows some frames having water and bubbles, which do not carry any useful visual information of mucosa. These frames need to be classified as non-informative. However, most IFF algorithms [5, 6] classify them as informative since they have clear edges and are in-focus. These types of frames are caused by water injection for cleaning purpose during the colonoscopy procedure, and need to be discarded from the further processing. We define a frame as water or bubble frame if more than 50% of the frame is covered with water or bubble. We call the frames in Fig. 3(a-b) 'water' frames, and the ones in Fig. 3(c-d) 'bubble' frames for convenience. Based on our observation with 100 colonoscopy videos, the percentage of these frames varies from 5.6% to 20.7% and 9.7% on average. Accurately detecting and discarding water and bubble frames can improve the performance of the 'automated feedback system' mentioned earlier [3, 4].

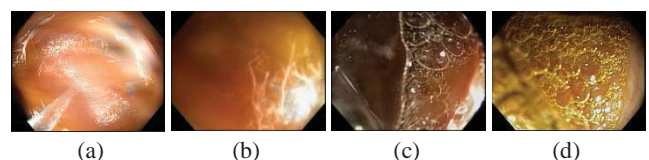


Fig. 3. Examples of Water/Bubble Frames: (a) and (b) Water frames, (c) and (d) Bubble frames.

In this paper, we propose a novel method for water and bubble frame detection based on image texture focusing on accumulation of pixel value differences. We compare it with other existing texture based algorithms in terms of accuracy



and execution time. To further reduce the execution time, we investigate different clustering methods. The proposed method performs very well in terms of accuracy and execution speed with or without clustering. More detailed explanation of accuracy and execution speed of the method is described in Section 4. Therefore, our main contribution is to propose a novel method which can detect water and bubble frames with very high accuracy in significantly less processing time.

The remainder of this paper is organized as follows. Related work is presented in Section 2. The proposed technique is described in Section 3. In Section 4, we discuss our experimental setup and results. Finally, Section 5 presents some concluding remarks.

## 2 Related work

To the best of our knowledge, water and bubble frame detection in colonoscopy videos has not been investigated before. The most closely related work is our previous work [5]. Recently, a new non-informative frame filtering method based on difference of Gaussian filtering has been proposed [6] but it has same limitation like ours in which very clear water and bubble frames can be classified as informative. The clustering of non-informative frames in GI endoscopy videos is proposed for manifold learning to create structured manifolds from complex endoscopic videos [7]. Color and texture based features are extracted to classify the colon status as either normal or abnormal using Principle Component Analysis to reduce the size of features [8]. Besides, there exist several texture detection techniques. The most commonly used ones are: Higher Order Local Auto Correlations (HLAC) [9], LBP (Local Binary Patterns) [10], Gabor filter banks [11], Leung-Malik filter banks [12], the traditional texture features (i.e., Contrast, Correlation, Energy, Homogeneity, etc.) based on Gray-Level Co-Occurrence Matrix (GLCM) [13], MPEG-7 texture features [14] as well as Discrete Fourier Transform (DFT) [15]. All of these methods are competitive in terms of accuracy but their execution speeds vary a lot. We present the evaluation method and results of most of these existing algorithms. Also, we will compare our proposed method and these existing methods in Section 4 in terms of both ‘with clustering’ and ‘without clustering’.

## 3 Methodology

The water and bubble textures are not uniform throughout the image resulting in a significant variations in the textures. Therefore, it is better to extract the features based on blocks rather than the entire image. We divide the images (720x480 pixels) into a number of blocks in which each block is 128x128 pixels. We experimented with various block sizes such as 32x32, 64x64, 128x128, and 256x256 pixels, and empirically determined the block size of 128x128 pixels to be optimal for capturing unique textures, and computationally efficient. For better capturing of the non-uniform textures, we allow an overlap for block division, which means one block overlaps 50% horizontally and vertically with its neighboring blocks. All extracted blocks will be preprocessed in the next step.

### 3.1 Preprocessing

Some of the blocks as seen in Fig. 4(a) and (b) need to be discarded from further processing. Fig. 4(a) shows a corner block with black borders consisting of many black pixels. Including these blocks in the feature extraction process may cause inconsistent results since black pixels in the borders are not from the colon. To filter out those blocks, first we separate red, green and blue channels from the original RGB color space, and normalize them so that the intensity value range for each channel becomes between 0 and 1. If all RGB channel values of a pixel are less than a threshold ( $BP_{Thld}$ ), it will be considered as a black pixel. If the black pixel percentage of a block is greater than a threshold ( $BBPP_{Thld}$ ), we discard it. We use RGB color space to avoid the processing overhead of color space conversion.

Some of the blocks like in Fig. 4(b) have very high uneven illumination which may provide incorrect characteristics of textures. The uneven illumination is characterized by calculating the standard deviation of the gray values of all the pixels in the block. We discard those blocks by thresholding with the standard deviation. If the standard deviation of a block is greater than a threshold ( $BSD_{Thld}$ ), we discard it. The actual values of the thresholds discussed above are summarized in Table 1, which are determined experimentally. The threshold values are determined one at a time using the entire training images. Once the block filtering is done, each remaining block such as shown in Fig. 4(c) is converted into a grayscale block for further processing. To make the grayscale properties consistent throughout the procedure, we normalize the block by subtracting the minimum grayscale value from each pixel in the block.

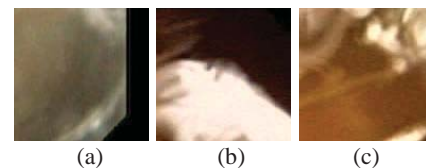


Fig. 4. Discarded blocks due to (a) black borders, (b) high standard deviation, and (c) Example of a block remained after block filtering.

TABLE 1. BLOCK FILTERING THRESHOLDS

<i>Parameters</i>	<i>Threshold values</i>
Black Border Pixel	$BP_{Thld} = 0.05$
Black Border Pixel Percent	$BBPP_{Thld} = 5.0\%$
Block Standard Deviation	$BSD_{Thld} = 0.25$

### 3.2 Proposed feature extraction method

We consider a window of 3x3 pixels, in which an average of the absolute differences between the center pixel ( $P_c$ ) and its eight neighbors ( $P_1 \sim P_8$ ) is calculated using (1), where  $n = 8$ .

$$DIFF(P_c) = \frac{1}{n} \sum_{k=1}^n |P_k - P_c|. \quad (1)$$

This process is repeated over the entire block. One reason why we consider 3x3 window is to provide a fair comparison with LBP where its best accuracy is achieved with this window size

[10]. A major difference between our new texture feature and others is that it considers not only patterns of pixel differences but also what center pixels are associated with them. A comparison between LBP and DIFF is shown in Fig. 5. As seen, LBP considers which pixel value is larger among a center and its neighbors, but does not consider how much larger whereas DIFF does retain the pixel value difference when it compares a center pixel value with its neighbors. For a block of 128x128 pixels, 15,876 (126x126) DIFF values are generated after excluding the border pixels. Minimum and maximum possible values of DIFF are 0 and 255, respectively. Also, possible values of  $P_c$  and  $P_k$  are between 0 and 255. Therefore, the number of occurrences of these values can be represented as a matrix (256x256) in which its columns and rows represent different DIFF and  $P_c$  values, respectively.

Now, the texture of a block is represented by 65,536 (256x256) numbers. We reduce this number by quantization. The quantization of the center pixel ( $P_c$ ) values is straightforward. Since it has 256 values, it can be quantized into any number by dividing by  $2^m$  ( $m = 1, 2, \dots, 8$ ). In our case we quantize it into 16 values ( $256/2^4$ ), 8 value ( $256/2^5$ ), 4 value ( $256/2^6$ ) and so on. The quantization reduces the size of the feature vector and thereby accelerates the performance whereas too much quantization may generate unreliable features depending on the nature of textures in the images.

DIFF values generated by (1) for the blocks of our colonoscopy images are typically less than 50, and mostly less than 10 based on the observation of our entire training images. In fact, the first 10 DIFF values (i.e., the first 10 bins in the histogram) represent more than 95% DIFF values of the entire block for all image types. The reason is that the pixel value differences in a 3x3 window are very small since neighboring pixels are very similar. We consider some combinations of quantized center pixel ( $P_c$ ) values with quantized DIFF values as new features such as 'DIFF\_2\_10' with 2 center pixel ( $P_c$ ) values and 10 DIFF values. Similarly, 'DIFF\_16\_10' with 16 center pixel ( $P_c$ ) values and 10 DIFF values, 'DIFF\_1\_10' with one center pixel ( $P_c$ ) value and 10 DIFF values, and so on. The selection of DIFF values and center values for a particular feature is dependent on the characteristics of textures in the blocks.

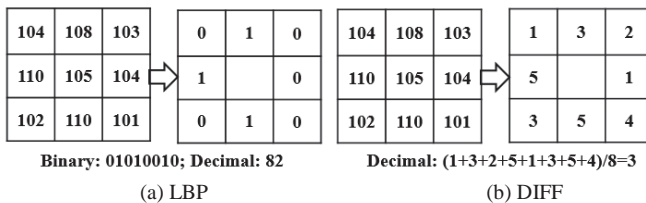


Fig. 5. Comparison of feature extraction between LBP and DIFF.

### 3.3 Evaluation method

Evaluation method has mainly two phases: Training and Testing. For Training, each input image is divided into a number of blocks, and the block filtering and normalization are applied as discussed above. A selected feature is computed for all blocks, and it is used to train a KNN (k-nearest neighbors)

classifier [16] with  $k=1$ . We experimented with different values of  $k$ , but found  $k=1$  giving best results for our dataset. We also tested other classifiers such as CART (Classification and Regression Tree) Decision Tree and SVM (Support Vector Machine) with linear kernel in MATLAB. Their comparison results will be discussed later. For Testing, a test image is divided into the same number of blocks with the same block size used in Training. The same block filtering and normalization as used in Training are applied to all blocks in the test image. Using the trained KNN classifier, we determine for each block to which type it belongs. Lastly, we calculate the probability of each type i.e. water/bubble or normal by dividing the detected number of blocks for each type by the total number of blocks processed for that image. If the test image has at least 50% water/bubble blocks, it is classified as a water/bubble frame. Otherwise, it is classified as normal frame. The rationale behind 50% threshold is that the colon mucosa is mostly hidden in an image covered with water/bubble by more than half. These types of frames will negatively affect the feedback system.

## 4 Experiments

All experiments were conducted on a Windows 7 64-bit PC with Intel i7 2.8GHZ processor and 6GB RAM using MATLAB R2014a. We present our results using commonly used performance metrics [17]: Recall (or Sensitivity) (R), Specificity (S), Precision (P), and Accuracy (A). They are based on Table 2 of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). Precision (P) computed as  $TP/(TP+FP)$  is the ratio of correctly classified positive instances from the predicted positives. Recall (or Sensitivity) (R) computed as  $TP/(TP+FN)$  is the ratio of correctly classified positive instances. Specificity (S) computed as  $TN/(TN+FP)$  is the ratio of correctly classified negative instances. The accuracy (A) computed as  $(TP+TN)/(TP+FP+FN+TN)$  is the ratio of correctly classified instances. The number of images and blocks used for training and testing are summarized in Table 3.

TABLE 2. EVALUATION METRICS

Actual	Predicted	
	Water/Bubble	Normal
Water/Bubble	TP	FN
Normal	FP	TN

TABLE 3. DESCRIPTION OF NUMBERS OF IMAGES AND BLOCKS USED

Type	Training		Testing	
	Image	Block	Image	Block
Water + Bubble	588	22,049	288	10,522
Normal	599	21,296	284	10,456
Total	1,187	43,345	572	20,978

### 4.1 Existing feature extraction

We compare several existing texture feature extraction methods such as Higher Order Local Auto Correlations (HLAC) [9], three versions (LBP59, LBP10, and Local

variance method (LOCAL\_VAR)) from the original LBP [10], Gabor filter banks (GABOR) [11], Leung-Malik filter banks (LM) [12], the traditional textures (Contrast, Correlation, Energy, and Homogeneity) based on Gray-Level Co-Occurrence Matrix (GLCM) [13], MPEG-7 based texture features (MPEG-7\_HTD (Homogeneous Texture Descriptor) and MPEG-7\_EHD (Edge Histogram Descriptor)) [14], and Discrete Fourier Transform (DFT) [15]. All of these feature extraction methods were implemented in MATLAB except the MPEG-7 descriptors which were implemented in C.

Higher Order Local Auto Correlations (HLAC) [9] is evaluated where the primitive features are obtained from the binary image by computing the sums of the products of the gray scale values of the corresponding pixels with 25 local 3x3 masks resulting in a 25 bin feature vector.

LBP (Local Binary Patterns) is a widely used method that describes the local texture patterns. Although it can be generalized to any size and any neighbors, we only focused on a 3x3 pixel neighborhood because it provides better accuracy according to [10], and it is computationally less expensive than larger neighborhoods. LBP59 contains 59 bins where the first 58 bins are 58 uniform patterns, and the last bin is everything else. Similarly, LBP10 is obtained as described in [10], and is called rotation invariant uniform LBP. It contains 10 bins where the first 9 bins contain the 9 rotation invariant uniform patterns, and the last bin contains all remaining 'non-uniform' patterns. The local variance method (LOCAL\_VAR) is obtained as described in [10], which takes the mean of 8 pixel neighbors, and subtracts the mean from each of those neighbors. For our experiment, we set the number of bins to 256 since larger bin size takes longer processing times.

We also considered Gabor filter banks with 80-bin feature vectors (GABOR) [11]. First, we obtain 40 different Gabor filters (5 scales and 8 orientations), and convolute each filter with the input block to get 40 different response matrices. We obtain Local Energy and Mean Amplitude by using the response matrices. Local Energy is calculated by summing up the squared values of a response matrix. Similarly, Mean Amplitude is calculated by summing up the absolute values of a response matrix. We included another popular filter bank called Leung-Malik (LM) [12] which is a multi-set, multi-orientation filter bank with 48 filters. It consists of first and second derivatives of Gaussians at 6 orientations and 3 scales making a total of 36 filters, 8 Laplacian of Gaussian (LOG) filters, and 4 Gaussian filters. Similar local energy and mean amplitude as in GABOR feature are computed to make a 96 bin feature vector.

Traditional texture features based on Gray-Level Co-Occurrence Matrix (GLCM) [13], which show the relationships between adjacent pixels are also included for the comparisons. Here its feature vector consists of four texture features (Contrast, Correlation, Energy, and Homogeneity).

For MPEG-7 based texture features, we have used Homogeneous Texture Descriptor (HTD) and Edge Histogram Descriptor (EHD) [14]. MPEG7\_HTD is composed of a 62 bin feature vector. The first two are the mean and the standard

deviation of the image block. The rest are the energy and the energy deviation of the Gabor filtered responses. MPEG7\_EHD represents local edge distributions in the image block by 80 bin feature vector.

We have explored Discrete Fourier Transform (DFT) [15] based feature as well. First, we get DFT of the input block using Fast Fourier Transform (FFT) algorithm [15]. To reduce the feature vector size, we take the mean and standard deviation of each row of the resultant block, and use them as features. In this way, a block is represented by a 256 bin feature vector with 128 means and 128 standard deviations.

## 4.2 Evaluation without clustering

We evaluated a total of 15 features including 5 different versions of our DIFF features (DIFF\_1\_10, DIFF\_1\_50, DIFF\_2\_10, DIFF\_8\_10, and DIFF\_16\_10). Table 4 shows the results in terms of precision, recall, specificity, and accuracy for both image and block levels. Most feature methods are providing decent (i.e., 90-95%) image level accuracies. Also, there is little difference in the block level accuracies. We observed that almost all of our feature methods (DIFF\_1\_50, DIFF\_2\_10, DIFF\_8\_10, and DIFF\_16\_10) performed on par with popular existing methods. We will discuss their performances with clustering and their computation costs later.

## 4.3 Evaluation with clustering

For more efficient and faster computing, we consider clustering of the blocks. To provide accurate detection of water and bubble frames, a huge number of blocks in the training set need to be compared with the blocks in an unseen image. By the clustering, we can reduce the number of comparisons, which impacts the execution speed. A cluster has hundreds of feature vectors generated from hundreds of blocks. Instead of comparing with these hundreds of vectors, we can compare with one vector which is its centroid (i.e., mean). For the clustering purpose, we use K-means, K-medoids and Fuzzy C-means clustering [16]. But, first we need to find an optimal number of clusters. To do that, we use the Elbow method which is simple but effective [18] where Within Cluster Sum of Squares (WCSS) is observed for different number of clusters. We ran K-means clustering for  $k = 10, 20, 30, \dots, K_{\max}$ , where  $K_{\max}$  equals 500 in our case, and the WCSS value is computed for each  $k$ . Our goal is to find the minimum value of  $k$  without sacrificing the accuracy of the classification.

Then, we plot clusters ( $k$ ) versus WCSS values. The optimal number of clusters is estimated by looking for  $k$  for which WCSS is not decreasing rapidly. Fig. 6 shows the plot for water/bubble blocks as well as normal blocks in which DIFF\_2\_10 feature is used for computing WCSS values. The plot was obtained based on all of the training blocks for both water/bubble and normal images as listed in Table 3. As seen in the plot, after the  $k$  value around 50, the WCSS values do not decrease rapidly. And, after the  $k$  value around 300, the WCSS values change very slowly, which makes the graph almost flat. So, we can see that an optimal  $k$  value can be in the range from 50 to 300. Next, we find the optimal  $k$  from this range.

TABLE 4. IMAGE AND BLOCK LEVEL ACCURACY WITHOUT CLUSTERING AND WITH CLUSTERING (INSIDE PARENTHESIS) (UNIT:%)

Features	Image				Block			
	Precision	Recall	Specificity	Accuracy	Precision	Recall	Specificity	Accuracy
DIFF_1_10	91.1 (90.8)	88.5 (85.4)	91.2 (91.2)	89.9 (88.3)	69.9 (70.7)	67.9 (66.7)	70.5 (72.2)	69.2 (69.4)
DIFF_1_50	91.0 (88.9)	89.9 (88.9)	90.8 (88.7)	90.4 (88.8)	70.0 (70.1)	67.9 (66.7)	70.8 (71.4)	69.3 (69.0)
DIFF_2_10	95.8 (97.0)	90.3 (89.9)	96.1 (97.1)	<b>93.2 (93.5)</b>	76.3 (75.5)	70.1 (69.4)	78.1 (77.4)	74.1 (73.4)
DIFF_8_10	96.0 (96.2)	91.7 (87.1)	96.1 (96.5)	93.9 (91.8)	77.1 (75.6)	72.0 (68.1)	78.4 (78.2)	75.2 (73.1)
DIFF_16_10	95.4 (95.1)	92.7 (87.5)	95.4 (95.4)	94.1 (91.4)	75.7 (74.7)	72.1 (69.5)	76.8 (76.4)	74.4 (72.9)
LBP10	88.8 (78.2)	91.3 (90.9)	88.4 (74.3)	89.9 (82.7)	72.8 (72.5)	72.7 (81.5)	72.7 (68.9)	72.7 (75.2)
LBP59	93.1 (90.5)	93.7 (89.6)	92.9 (90.5)	93.4 (90.0)	76.4 (79.2)	75.4 (76.1)	76.6 (79.9)	75.9 (78.0)
HLAC	91.6 (87.7)	94.4 (84.0)	91.2 (88.0)	92.8 (86.0)	69.9 (66.5)	72.9 (65.7)	68.4 (66.8)	70.7 (66.2)
GLCM	86.3 (78.9)	85.4 (87.1)	86.2 (76.4)	85.8 (81.8)	63.6 (62.3)	62.8 (63.2)	63.9 (61.6)	63.3 (62.4)
LOCAL_VAR	91.1 (89.2)	82.3 (77.4)	91.9 (90.5)	87.1 (83.9)	69.2 (70.5)	65.0 (65.2)	70.9 (72.6)	67.9 (68.9)
GABOR	94.1 (97.7)	93.7 (45.1)	94.0 (98.9)	93.5 (71.8)	74.0 (87.6)	74.7 (47.5)	73.7 (93.3)	74.2 (70.3)
LM	93.4 (95.1)	93.7 (93.7)	93.3 (95.1)	93.5 (94.4)	73.4 (71.7)	74.2 (74.4)	73.0 (70.5)	73.6 (72.5)
DFT	95.9 (96.6)	90.3 (78.1)	96.1 (97.2)	93.2 (87.6)	70.5 (74.4)	67.2 (61.5)	71.7 (78.7)	69.5 (70.0)
MPEG7_HTD	96.7 (81.5)	93.1 (95.1)	96.8 (78.2)	94.9 (86.7)	77.8 (67.2)	74.6 (75.5)	78.6 (62.9)	76.6 (69.2)
MPEG7_EHD	77.8 (66.2)	92.7 (95.8)	73.2 (50.3)	83.0 (73.2)	67.4 (60.6)	78.4 (74.2)	61.8 (51.3)	70.1 (62.8)

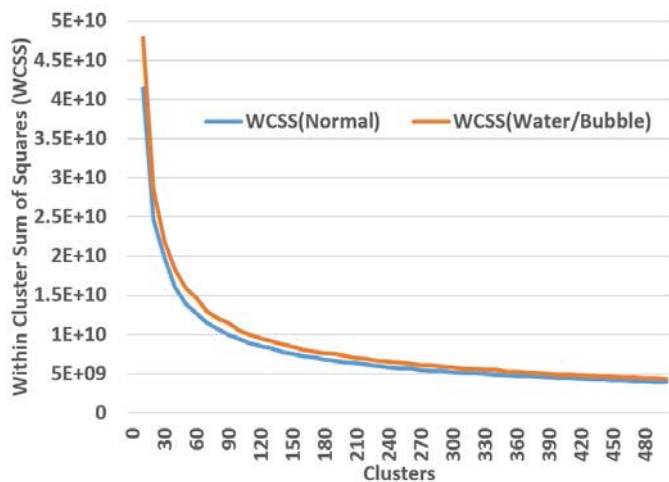


Fig. 6. Optimal cluster estimation using Within Cluster Sum of Squares (WCSS).

We tested several different k values. Fig. 7 and 8 show the image and block level accuracies when K-means clustering is used with different cluster sizes and with DIFF\_2\_10 as feature. It can be seen that the best image and block level accuracies are achieved at the cluster size of around 200 which falls in the estimated range by WCSS plot. The numbers inside the parentheses in Table 4 show the results in terms of precision, recall, specificity, and accuracy for both image and block levels with clustering of size 200. As seen, the performances are degraded for the most of the features when compared with those without clustering. DIFF\_1\_10, DIFF\_2\_10, DIFF\_8\_10, DIFF\_16\_10, LBP59, and LM are still good (i.e., better than 90%). We claim that DIFF\_2\_10 is our choice since it is faster than the others. A speed comparison of feature extraction of various features with and without clustering will be discussed later.

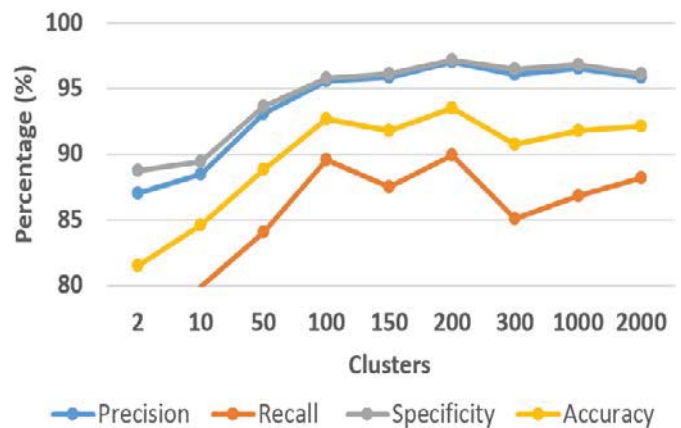


Fig. 7. Image level accuracy for different numbers of clusters using DIFF\_2\_10 feature.

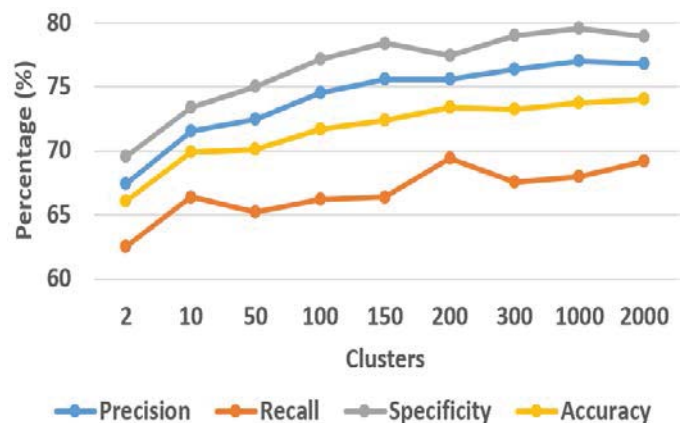


Fig. 8. Block level accuracy for different numbers of clusters using DIFF\_2\_10 feature.

We also evaluated our best performing feature DIFF\_2\_10 using different clustering algorithms and classifiers. We set the number of cluster to 200 as before and cluster our training blocks using K-medoids and Fuzzy C-Means as well as K-means clustering algorithms [16]. For the classification we chose SVM and CART Decision Tree to compare with KNN [16]. Fig. 9 and 10 show the results of a total of nine combinations of the three clustering algorithms and three classifiers for the image and block levels, respectively. As seen, KNN with K-means clustering is the best among all in terms of accuracy. We observed that k-means and decision tree combination gives best recall percentage but it is not as good in terms of other metrics. We excluded the classification results of the other features because of space limitation.

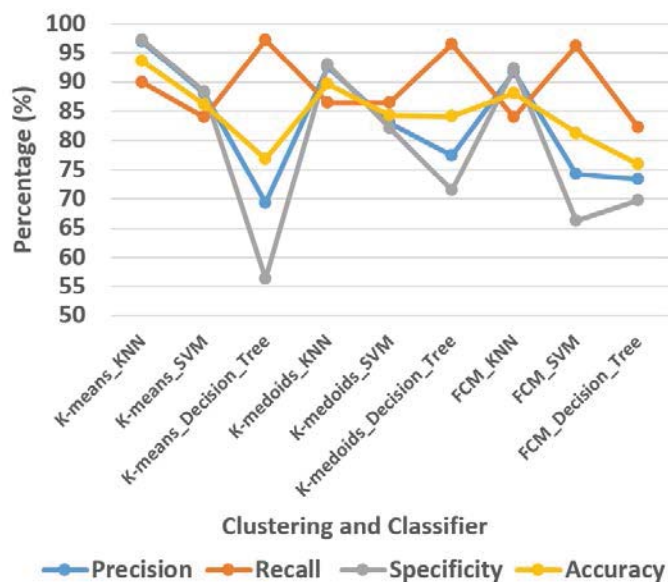


Fig. 9. DIFF\_2\_10 image level accuracies with 3 different clustering algorithms and 3 classifiers.

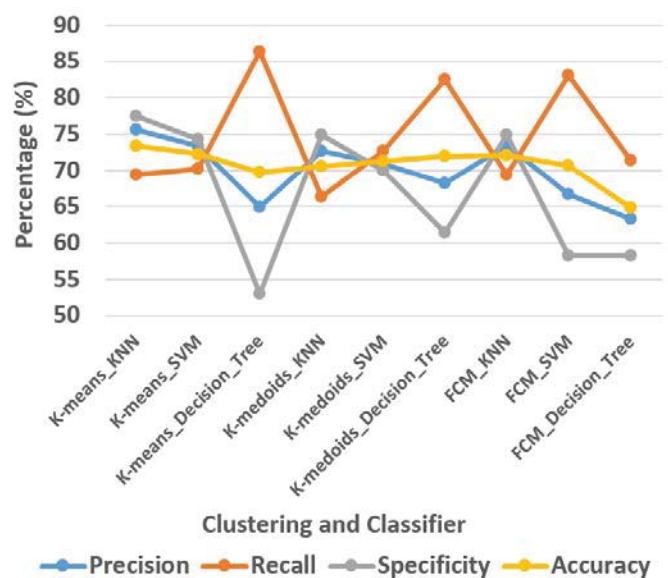


Fig. 10. DIFF\_2\_10 block level accuracies with 3 different clustering algorithms and 3 classifiers.

#### 4.4 Execution speed comparison

Computation cost is really important in a colonoscopy video processing system since a very large number of frames need to be evaluated. We compare the computation costs of some of the better performing features. Tables 5 and 6 show the results of the total computation costs for entire images of training and testing listed in Table 3 for both ‘without clustering’ and ‘with clustering’. As seen, our DIFF based features are more than 2 times faster than the others for the training phase in both with and without clustering evaluations. For the testing phase, our best performing feature DIFF\_2\_10 is significantly faster than all other similarly performing features. For example – per frame testing cost for DIFF\_2\_10 is 746.9/572 (these numbers are from Tables 3 and 5) = 1.3 seconds. Since all the implementations are done in MATLAB, the cost can be reduced significantly once implemented in C/C++. As mentioned earlier, the main reason for clustering is to reduce the number of comparisons in the testing phase thereby reducing the computation cost. We observed that the computation cost improves dramatically in the testing phase for feature methods with a larger feature vector size like DIFF\_16\_10 (160 bin) and DFT (256 bin). For example – the testing time of DIFF\_16\_10 is reduced more than 2 times with clustering. Even without considering the one-time cost like clustering and training, classification using our DIFF based features is significantly faster than that using the other feature

TABLE 5. EXECUTION TIME WITHOUT CLUSTERING (UNIT: SECONDS)

Features	Training	Testing	Total
DIFF_1_10	251.4	668.8	920.2
DIFF_2_10	<b>278.6</b>	<b>746.9</b>	<b>1,025.5</b>
DIFF_8_10	278.0	1,101.1	1,379.1
DIFF_16_10	283.1	1,562.9	1,846.0
LBP10	965.9	1,061.3	2,027.2
LBP59	857.3	1,251.6	2,108.9
HLAC	1,278.9	1,272.2	2,551.1
DFT	677.2	2,311.9	2,989.1
MPEG7_HTD	1,842.1	2,061.3	3,903.4
LM	8,565.0	4,941.6	13,506.6
GABOR	16,713.2	13,407.6	30,120.8

TABLE 6. EXECUTION TIME WITH CLUSTERING (UNIT: SECONDS)

Features	Clustering	Training	Testing	Total
DIFF_1_10	775.6	244.0	624.5	1,644.1
DIFF_2_10	<b>794.7</b>	<b>271.5</b>	<b>642.4</b>	<b>1,708.6</b>
DIFF_8_10	1,167.0	327.3	674.5	2,168.8
DIFF_16_10	2,513.5	370.5	675.9	3,559.9
DFT	4,861.3	749.8	846.4	6,457.5
LBP10	1,254.2	916.3	959.9	3,130.4
LBP59	1,287.5	827.9	979.6	3,095.0
HLAC	1,763.7	1,281.2	1,172.4	4,215.3
MPEG7_HTD	4,003.2	2,109.9	1,605.0	7,718.1
LM	9,477.5	8,825.1	3,727.2	22,029.8
GABOR	18,408.8	17,142.5	8,532.2	44,083.5

extraction methods. This shows that our DIFF based features are computationally efficient without sacrificing accuracy.

## 5 Conclusion

To improve quality of colonoscopy, we have been investigating an 'automated feedback system' which informs the endoscopist of possible sub-optimal inspection during the procedure. One of the basic steps of this system is to distinguish non-informative frames from informative ones. Existing methods for this cannot classify water/bubble frames (which do not carry any useful visual information of mucosa) as non-informative frames since they focus on image clarity not image semantic. To consider image semantic, we propose a novel image texture feature based on accumulation of pixel differences, which can detect water and bubble frames with very high accuracy and significantly less processing time. To reduce processing time even more, we employ clustering which can reduce the number of time-consuming comparisons. The experimental results show the proposed feature can achieve more than 93% overall accuracy in almost half of the time existing methods take.

## 6 Acknowledgement

This work is partially supported by EndoMetric Corp. Johnny Wong, Wallapak Tavanapong, and JungHwan Oh hold positions at EndoMetric Corporation, Ames, IA50014, USA, a for profit company that markets endoscopy-related software. De Groen is the medical advisor of EndoMetric.

## 7 References

- [1] American Cancer Society, 2015. "Colorectal Cancer Facts and Figures".
- [2] C. D. Johnson, J. G. Fletcher, R. L. MacCarty, et al, "Effect of slice thickness and primary 2D versus 3D virtual dissection on colorectal lesion detection at CT colonography in 452 asymptomatic adults," *American Journal of Roentgenology*, 189(3), pp. 672-680, 2007.
- [3] J. Oh, S. Hwang, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. C. de Groen, "Measuring Objective Quality of Colonoscopy," *IEEE Transaction in BioMedical Engineering*, 56(9), pp. 2190-2196, 2009.
- [4] S. Stanek, W. Tavanapong, J. Wong, J. Oh, R. D. Nawarathna, and J. Muthukudage, and P. C. de Groen, "SAPPHIRE: A Toolkit for Building Stream Programs for Medical Video Analysis," *Computer Methods and Programs in Biomedicine*, 112(3), pp. 407-421, Dec. 2013.
- [5] V. P. Karri, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen, "Effective and Accelerated Informative Frame Filtering in Colonoscopy Videos using Graphics Processing Unit," *Proceeding of International Conference on Bio-inspired Systems and Signal Processing*, pp.119-124, Jan. 2011.
- [6] B. Munzer, K. Schoeffmann, and L. Boszormenyi, "Relevance Segmentation of Laparoscopic Videos," 2013 *IEEE International Symposium on Multimedia (ISM)*, pp.84-91, Dec. 2013.
- [7] S. Atasoy, D. Mateus, J. Lallemand, A. Meining, C. Z. Yang, and N. Navab, "Endoscopic Video Manifolds," *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 437-445, 2010.
- [8] M. P. Tjoa, and S. M. Krishnan, "Feature extractions for the analysis of colon status from the endoscopic images," *Biomedical Engineering Online*, vol. 2, no. 9, pp.1 -17, 2003.
- [9] T. Kurita, N. Otsu, and T. Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," *11th IAPR International Conference in Pattern Recognition*, vol. II. pp. 213-216, 1992.
- [10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971-987, 2002.
- [11] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, "Identification Using Encrypted Biometrics," in *Computer Analysis of Images and Patterns*. vol. 8048, pp. 440-448, 2013.
- [12] T. Leung, and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, 43(1) pp. 29-44, Jun. 2001.
- [13] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, pp. 610-621, 1973.
- [14] M. Bastan, H. Cam, U. Gudukbay, and O. Ulusoy, "BiVideo-7: An MPEG-7 Compatible Video Indexing and Retrieval System," *IEEE MultiMedia*, vol. 17, no. 3, pp. 62-73, Sep. 2010.
- [15] P. Duhamel, and M. Vetterli, "Fast Fourier Transforms: A Tutorial Review and a State of the Art," *Signal Processing*, vol. 19, pp. 259-299, Apr. 1990.
- [16] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," *John Wiley and Sons*, pp. 517-581, 2001.
- [17] J. Han, and M. Kamber, "Data Mining: Concepts and Techniques," *Morgan Kaufmann Publishers*, 2001.
- [18] D. J. Ketchen, and C. L. Shook, "The application of cluster analysis in Strategic Management Research: An analysis and critique," *Strategic Management Journal* 17 (6): 441-458, 1996.

# Non-Invasive Heart Murmur Detection via a Force Sensing-Based System and SVM Classification

Samuel W. Rud, Nicholas St. Jacques, Aaron D. Vant and Jiann-Shiou Yang  
Department of Electrical Engineering, University of Minnesota, Duluth, MN 55812, USA

**Abstract** — Noninvasive heart murmur detection is the process of diagnosing a patient's heart condition without intrusion into the body. By providing a physician with effective and economical tools to help aid in the diagnosis could decrease misdiagnosis rates. This paper presents a study of detecting low frequency vibrations on the human chest via a FlexiForce force sensing technique and then correlate them to cardiac conditions via the Support Vector Machine (SVM) classification technique. The force sensor-based system is implemented and, with the consent of the University of Minnesota Institutional Review Board (IRB), the system was tested through clinical trials. A Support Vector Machine (SVM) learning algorithm is then used to train, tune, and classify signals. Our preliminary study indicates that a SVM is able to distinguish signals between normal and abnormal cardiac conditions.

**Keywords:** Heart murmur; support vector machine, classification.

## I. INTRODUCTION

HEART murmurs are sounds caused by turbulent blood flow through a heart's valve. Turbulence is present when the flow across the valve is excessive for the area of the open valve. It can be due to normal flow across a diseased valve, abnormally high flow across a normal valve, or a combination. Murmurs are classified as systolic, diastolic, or innocent to describe the location of the turbulent blood flow in the heart [1]. Noninvasive heart murmur detection is the process of diagnosing a patient's heart condition without intrusion into the body. The process has evolved in many different directions stemming from listening to the human chest with a stethoscope to using computers to detect heart conditions [2-4]. The simplest and most formal way of diagnosing heart murmurs is via a primary care physician's judgment on what he or she heard. However, primary care physicians accurately associate a murmur diagnosis with its pathology as little as 20% of the time. By providing a physician with effective and economical tools to help aid in the diagnosis could decrease misdiagnosis rates. The aim of this paper is to focus on detecting low frequency vibrations on the human chest using a force sensing technique and then correlate them to cardiac conditions via the Support Vector Machine (SVM) classification technique. We mainly target vibrations with frequencies ranging from

10 to 150 Hz. The SVM classification is used to devise a computationally efficient way of learning "good" separating hyperplanes in a high dimensional feature space. Since the invention of SVMs by Vapnik [5, 6], there have been intensive studies on the SVM for classification and regression (e.g., [7, 8]). Recently, the application of SVM to various research fields have also been reported in the literature (e.g., the intelligent transportation systems (ITS) [9, 10]). This paper will describe the methodology and results for detecting and classifying heart murmurs.

## II. METHODS

### A. Device, Circuits and Software

The primary sensor device used in this study is the TekScan A201 FlexiForce sensor. It uses an aluminum dye that inversely changes resistance when force is applied on the dye. The sensor's range in force applied is approximately 0-1 pound, corresponding to a resistance range of 5 M $\Omega$  to 1 k $\Omega$  (<http://www.tekscan.com/>). The FlexiForce sensor device shown in Fig. 1 is used to measure force between two objects.

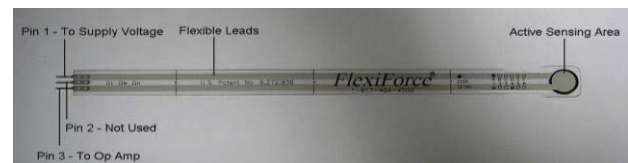


Fig. 1 TekScan A201 FlexiForce Sensor

The hardware part includes a drive circuit utilizing Texas Instrument's OPA228 operational amplifier. The FlexiForce electrical circuit is composed of an inverting operational amplifier with the FlexiForce sensor acting as a variable resistor as shown in Fig. 2. The circuit consists of (1) a Texas Instruments OPA4228N op-amp, (2) a 20 k $\Omega$  resistor, and (3) a FlexiForce 1 pound sensor. The circuit measures the load or force on the FlexiForce sensor as a voltage.

An EKG circuitry, shown in Fig. 3, was designed to electrically isolate the patient (<http://www.picotech.com/>). The design keeps the leakage current to a minimum to prevent electrical shock. The heart would seize if enough current traversed it, therefore, electrical isolation is needed

to be ensured for this reason. The sensor's signal is amplified via a non-inverting amplifier with an OPA4228N op-amp. In addition, several analog high-pass filters are used in the system to eliminate any DC offset produced by the sensors. This allows for optimizing the range of frequencies existing in a system. The analog filters have a high enough corner frequency to eliminate any breathing cycles produced by a subject, but also low enough not to interrupt the target frequencies of the signal. The detailed information about our designed 1<sup>st</sup> and 2<sup>nd</sup> order high-pass filters, with corner frequencies at 4.8214 Hertz (Hz) and 7.4881 Hz, and their frequency responses are available upon request.

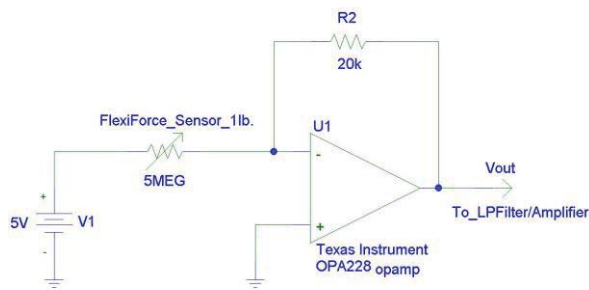


Fig. 2 FlexiForce drive circuit.

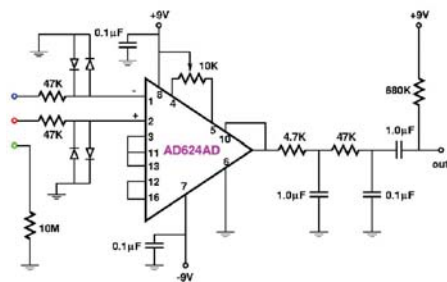


Fig. 3 EKG schematic.

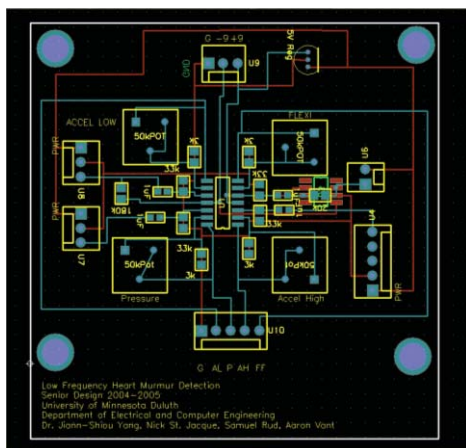


Fig. 4 Printed circuit board.

The PCB, shown in Fig. 4, is composed of four first-order passive high-pass filters with some additional hardware and circuitry to deliver the driving signal to the appropriate sensor system. This board was designed and implemented using Pad2Pad software. It creates a low noise situation that a solderless bread board introduces. This allows for a smaller signal to noise ratio thus enabling a cleaner signal to be recorded. Since the signals that are being recorded are generally low amplitude, large amount of amplification is needed. In doing so, noise is introduced in the system and this PCB helps to minimize the noise effect.

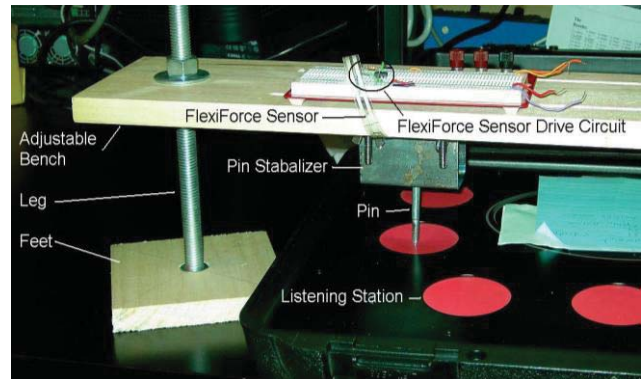


Fig. 5 Bench and FlexiForce circuit.

The bench assembly in Fig. 5 provides a static surface for the FlexiForce system to keep the sensor under constant load. It also provides a firm surface to assist the sensor in creating a linear resistance characteristic. The bench mainly consists of a piece of poplar lumber measuring approximately two feet in length. This number is able to accommodate a human body or other equipment when the bench is placed over it. Sheet metal was bent to act as a guide for the vibration transmitter which is shown in Fig. 5 labeled “pin”. The legs, which act as risers, are constructed by the shaft of a bolt. This allows alterations to be made by spinning a nut on the riser for fine tuning of the system. The circuit that drives the FlexiForce sensor is able to rest on the top of the board as shown in Fig. 5. The signal generated by these circuits is processed by a Measurement Computing PCI-DAS6036 data acquisition card which interfaces with a custom Matlab software application to record the signals. The sensor device observes vibrations from the human chest with interest in the lower range of frequencies of 10 to 150 Hz. The signals are read into the computer through a data acquisition system and recorded. The software application dubbed “MurmurPro”, a graphical user interface (GUI) package that combines data acquisition, signal analysis, and signal detection into one package, allows us to view recorded signals, analyze and process the recorded signals, and later classify the signals using a Support Vector Machine (SVM) learning algorithm.



### B. Testing Procedure

The testing environment, with consent of the University of Minnesota Institutional Review Board's (IRB) Human Subjects Committee, was human patient testing (or clinical trials). The FlexiForce system acquired voltage signals indirectly through a vibration transmitter placed on the surface of the subwoofer. The vibration transmitter experiences vibrations from the speaker within and transfers them to the FlexiForce sensor which is positioned between the vibration transmitter and the adjustable bench. The circuitry for the FlexiForce sensor system existed on top of the height adjustable bench.

### C. Support Vector Machine Training and Testing

The SVM is used to create an optimized boundary that separates between normal and abnormal cardiac conditions. It is a multi-step process as shown in Fig. 6 that includes signal processing, two-dimensional signal transformation, SVM tuning and optimization, and SVM Training. Signal processing entails filtering the signal to eliminate excessive noise and cropping the signal to a fixed number of heart beats. This is done to standardize each signal such that a two-dimensional representation can be achieved. The next process is then to tune the SVM. SVM tuning is a function provided by the Matlab's LS-SVM toolbox [11] that automatically tunes the SVM. It automatically tunes the hyper-parameters of the model with respect to the given performance measure. Hyper-parameters adjust the SVM model such that a desired plane can be achieved. The performance measure used in the software is called "grid search", which finds the minimum cost/error of the model based on a two-dimensional hyper-parameter search. If the user of MurmurPro chooses to optimize the system, a Bayesian optimization algorithm is provided by LS-SVM. This will further tune the hyper-parameters based on Bayesian statistical methods [11]. Finally, the SVM is trained using the tuned hyper-parameters and the given data. This yields a two-dimensional visual representation of the SVM with each signal's data points and the optimized boundary plotted.

## III. RESULTS

### A. Clinical Trials

The signals come from multiple patients in which the signals were categorized by the physician's diagnosis as normal or abnormal and pathological or non-pathological.

Fig. 7 shows the results obtained while testing a normal heart recording played through the subwoofer. This figure contains two plots. The first plot is the filtered time-series (TS) plot (voltage (V) vs. time (s)) while the plot following is the frequency response (FR) (magnitude (dB) vs. frequency (f)) plot of the filtered phonocardiogram. The

frequency response plots was created using the Fast Fourier Transform (FFT) method. In the TS plot, S1 and S2 are distinguishable and separable while the FR plot shows a frequency range of 10 to 300 Hz. Note that S1 and S2 represent the timings of the sounds of a normal heart cycle, where S1 is the first sound and S2 is the second sound. For this study, it is assumed that the major and minor peaks in the signal occur in sync with S1 and S2 respectively.

Fig. 8 shows the results obtained while testing an aortic stenosis recording played through the subwoofer. Fig. 8 illustrates the presence of extra vibration after S1. The FFT shows relatively the same frequency range as the normal heart recording in Fig. 7. However, several spikes are present that are not in the normal heart recording.

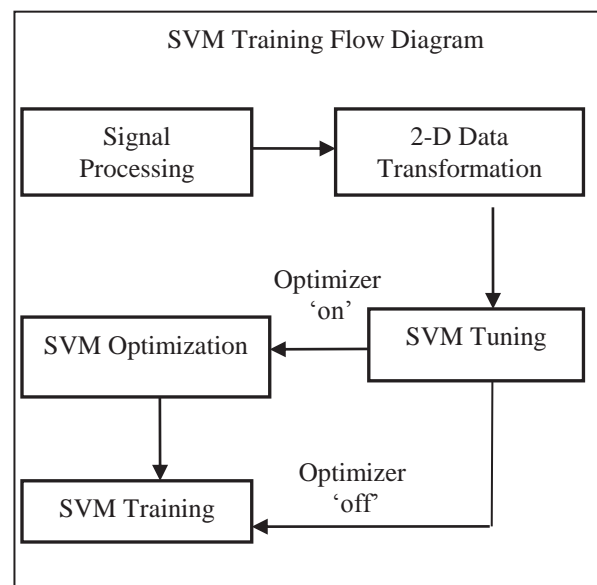


Fig. 6 Flow diagram of SVM training and tuning.

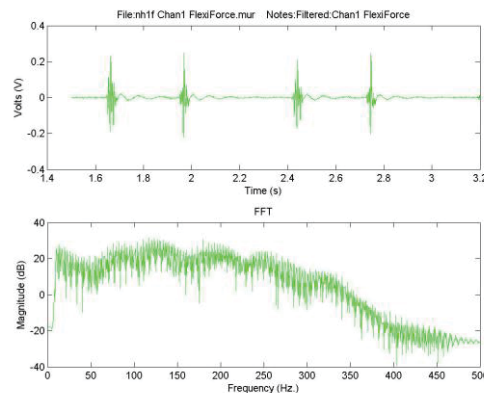


Fig. 7 Normal heart recorded by the FlexiForce sensor.

## B. SVM Classification

The signals from the device include the following heart conditions: (a) aortic stenosis; (b) Austin Flint murmur; (c) diastolic murmur; (d) friction rub; (e) mitral valve prolapse; (f) systolic murmur; and (g) normal heart. Each SVM was first trained with a training set and then tested with a testing set. The testing set did not contain any of the same signals as the training set. Also, heart rates of the signals ranged from 60 bpm to 80 bpm.

Fig 9 shows the trained SVM. Notice how normal heart signals tend to cluster together in a linear fashion, while the abnormal heart signals tend to be sporadic. This is due to the fact that normal heart signals should have a similar duration time and magnitude (depending on heart rate). The abnormal signals separate away from the normal signals, since the abnormal signals contain more vibrations giving a longer duration time and a larger magnitude.

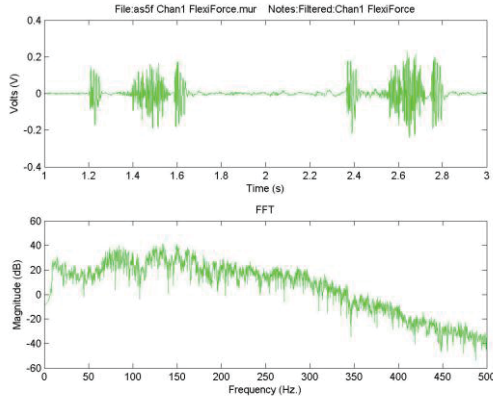


Fig. 8 Aortic Stenosis recorded by the FlexiForce sensor.

Fig. 10 shows the tested SVM. With 14 test cases, varying from normal to abnormal, only 2 misclassifications occurred resulting in 14% misclassification. The two test cases that were misclassified were one normal and one abnormal.

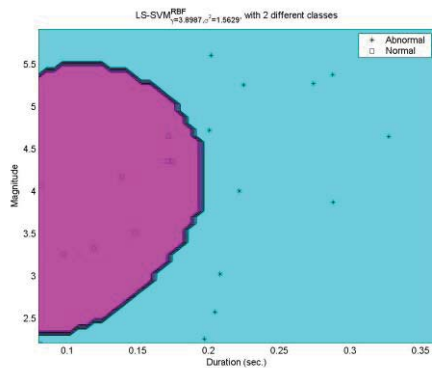


Fig. 9 Subwoofer FlexiForce trained SVM.

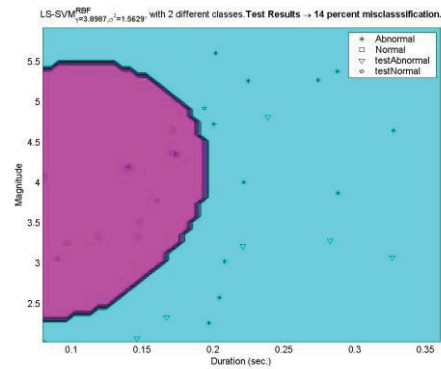


Fig. 10 Subwoofer FlexiForce tested SVM.

## IV. CONCLUSION

This paper focuses on the study of heart murmur detection and classification. The experimental setup via a FlexiForce sensor together with hardware and software interfaces was developed to detect low frequency vibrations from a human chest (clinical trials) in the targeted frequency range of 10 to 150 Hz. The purpose of patient testing is to provide reasonable evidence that this system can perform on humans.

The preliminary results show that the described technique has the capacity to capture distinguishable signals. During clinical trials, it was seen that the signal generated from a patient with a normal heart could be distinguished from a person with an abnormal heart via the FlexiForce sensor based system. Also, the SVM plots showed that it is possible to classify the signals. Unfortunately, the test data was limited to the amount of patients received. Also, there was no electro-cardiogram present at the clinical trials to establish the timings of the heart. This is due to the limits established by the IRB. Future developments of this study include improvements on the sensor system's design and implementation. Also, improvements on the SVM two-dimensional representation algorithm would further classify the signals more effectively. In addition, a real-time classification scheme could be devised to render an immediate diagnosis. Finally, extensive medical trials should be conducted to verify the sensor system and their respective classification accuracy rates.

## REFERENCES

- [1] O. Epstein, *et al*, *Clinical Examination*, New York: Gower Medical Publishing, 1992.
- [2] Watrous, *et. al*, "Computer-Assisted Detection of Systolic Murmurs Associated with Hypertrophy Cardiomyopathy," *Texas Heart Institute Journal*, vol. 31, no. 4, pp. 368.
- [3] K. Ejaz, *et. al*, "A Heart Murmur Detection System Using Spectrograms and Artificial Neural Networks," *Proc. IASTED Int'l Conf. Circuits, Signals, and Systems*, pp. 374-379, 2004.
- [4] N. Andrisevic, *et. al*, "Detection of Heart Murmurs Using Wavelet Analysis and Artificial Networks," *Journal of Biomechanical Engineering*, vol. 127, pp. 899-904, November 2005.
- [5] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York, NY: Springer, 1995.

- [6] V. N. Vapnik, "An Overview of Statistical Learning Theory," *IEEE Trans. Neural Networks*, vol. 10, pp. 988-999, September 1999.
- [7] K.R. Muller, A. J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. Vapnik, "Using Support vector Support Machines for Time Series Prediction," in *Advances in Kernel Methods*, pp. 242-253, Cambridge, MA: MIT Press, 1999.
- [8] S. R. Gunn, "Support Vector Machine for Classification and Regression," Technical Report, University of Southampton, Southampton, K.K., May 1998.
- [9] D. Gao, J. Zhou, and L. Xin, "SVM-based Detection of Moving Vehicles for Automatic Traffic Monitoring," *Proc. IEEE 4<sup>th</sup> Int'l Conf. Intelligent Transportation Systems*, pp. 745-749, 2001.
- [10] R. Reyna, A. Giralt, and D. Esteve, "Head Detection Inside Vehicles with a Modified SVM for Safer Airbags," *Proc. IEEE 4<sup>th</sup> Int'l Conf. Intelligent Transportation Systems*, pp. 500-504, 2001.
- [11] K. Pelckmans, et. al., LS-SVM Toolbox User's Guide, Version 1.4, Dept. Of Electrical Engineering, Katholieke Universiteit Leuven, October 2002.

# Moebius Strip Like Pathology: Mechanisms, Diagnosis, Treatment Correction

S. Kulishov<sup>1</sup>, O. Iakovenko<sup>2</sup>

<sup>1</sup> Internal medicine No 1, HMEIU "UMSA", Poltava, Ukraine; <sup>2</sup> Daily permanent establishment, V.P. Komissarenko Institute of Endocrinology and Metabolism AMS Ukraine Kyiv, Ukraine

**Abstract** *Proposed and tested an algorithm for diagnosis of Moebius strip like pathology, as prerequisite for treatment correction. The algorithm is reduced to initialization of study objects as a Moebius strip, in particular symptoms, syndromes, diseases, multimorbid states; clarification of investigation objects as non-orientable two-dimensional surface; cutting a Moebius strip like clinical data, variability to form two disjoint Moebius strips; clarification of chirality Moebius strip turn; the determination of Moebius strip like constituents as the object of research and their 3D representation; clarification of Moebius strips turn chiralities of constituents; the adoption of diagnostic and therapeutic solutions based on geometry of the pathogenetic and sanogenetic mechanisms. Thus, our algorithm may be basis for making diagnosis and treatment decision.*

**Keywords:** *Moebius strip like pathology, diagnosis*

## 1. Prerequisites of the application of Moebius strip like pathology mechanisms

It's known the Moebius strip like space orientation of depolarization processes were characterized by the change of supraventricular pacemaker on ventricular and inverse [1]. Moebius strip like arrhythmias in the patients with sinus node dysfunction were displayed as a combination of supraventricular and ventricular extrasystoles, fibrillation and flutter transformation from atria to ventricles [1].

The patients with complete atrioventricular block had the Moebius strip like changes of depolarization / repolarization geometry, as the alternation of proximal and distal ventricular rhythms [1].

The specifics of the geometry of depolarization and repolarization processes in the patients with full atrioventricular block or binodal syndrome may be considered in elaborating differential treatment programs to be used in microcomputers for implantable cardiac pacemakers [1].

Analysis of the cardiac depolarization / repolarization geometry may serve as additional criteria for sudden death prognosis [1].

Anxiety and depression can be a manifestation similar to the Moebius strip [2]. When comorbidity prevalence is over 50%, doubts about the validity of the diagnosis may be raised [2]. If comorbidity is so common, does this reflect a weakness of the diagnosis [2]?

Does it mean that the comorbidity itself should be a diagnosis? Or do the comorbid conditions reflect one underlying entity, with common etiology, expressed in two different phenotypes [2]? This might be the case with anxiety and depression. Actually, they are so clearly linked that the relationship resembles to many a Moebius strip [2]. Like in the Moebius strip, when looking at it, it is impossible to say where one edge ends and another begins [2].

Cutting along the middle of a Moebius strip makes a similar construct, more twisted, but larger and more easily examined and described [2]. Cutting along the middle of the new, larger shape results in two separate loops; each more twisted, and very closely intertwined, but nevertheless two distinct entities [2].

It remains to be seen whether continued research will reveal comorbid anxiety and depression to be two intertwined disorders or a single pathology with two different phenotypes [2].

The hope is that using various underlying endophenotypical tools will assist in unraveling the nature of the anxiety-depression [2].

Does depression exist without anxiety [3]? The affective dysregulation characterizing the comorbidity between anxiety and depression in a range of clinical populations is typical manifestation [3].

Anxiety is an integral and essential part of depression [4]. Moreover, anxiety symptoms should be considered a significant predictor of depression severity and the level of a patient's functional recovery, and can be utilized in choosing a treatment intervention [4].

Depression and anxiety disorders share a significant nonspecific component that encompasses general affective distress and other common symptoms [5]. General distress, physiological tension, and hyperarousal are more specific to anxiety, and melancholia and pervasive anhedonia are more specific to depression [5].

The purpose of this investigation was to determine Moebius strip like pathology as basis for making diagnosis and treatment decision.

## 2. The methodology of determination Moebius strip like pathology for making diagnosis and treatment decisions

Algorithm for diagnosis of Moebius strip like pathology, as prerequisite for treatment correction, is reduced to:

- Initialization of study objects as a Moebius strip like, in particular symptoms, syndromes, diseases, multimorbid states;
- Clarification of investigation clinical objects, variabilities as non-orientable two-dimensional surface, that are embedded in three-dimensional Euclidean space;
- Cutting a Moebius strip like objects of research in the middle to form two disjoint Moebius strips;
- Clarification of chirality Moebius strip turn;
- The determination of Moebius strip like constituents (symptoms, syndromes, diseases, multimorbid states) as the object of research and their 3D representation;
- Clarification of Moebius strips turn chiralities of constituents;
- The adoption of diagnostic and therapeutic solutions based on geometry of the pathogenetic and sanogenetic mechanisms

Anxiety and depression [2], pro-ischemic and anti-ischemic, reperfusion and anti-reperfusion, stress and anti-stress, hypoxic and antihypoxic, pro-oxidant and antioxidant, inflammatory and anti-inflammatory, dyslipidemic and anti-dyslipidemic, pro-arrhythmic and anti-arrhythmic, destructive and anti-destructive processes can be manifestations similar to the Moebius strip. When comorbidity, co-processes are over 50% prevalence, doubts about the validity of the diagnosis may be raised [2]. The comorbid conditions, co-processes may reflect same underlying entities, with common etiology, expressed in two or more different phenotypes [2]. Actually, they are so clearly linked that the relationship resembles to many a Moebius strip [2]. Like in the Moebius strip, when looking at it, it is impossible to say where one edge ends and another begins [2].

### 3. Implementation of Moebius strip like pathology making diagnosis and treatment decision

An example of Moebius strip like low-intensity inflammation syndrome diagnosis in the patients with coronary heart disease, arterial hypertension is presented lower. It is known that the leading role in the pathogenesis of coronary heart disease, hypertension belongs to disbalance between pro-inflammatory and anti-inflammatory cellular and plasma factors. Moebius strip like processing of inflammation [6,7,8,9,10,11] is represented as 2 variants:

1. Identification of multiple foci of pro-inflammatory factors' activation and the consumption of anti-inflammatory, leading to the widespread inflammatory, including autoimmune processes to chaperones;

2. Identification of multiple foci of anti-inflammatory factors' activation with the consumption of pro-inflammatory, leading to the widespread activity of anti-inflammatory interleukins and chaperones.

An example of diagnosis a composite plurality of pathological process as Moebius strip like in the patients with coronary heart disease may be by the next algorithm.

It is known that the basis of coronary artery disease are disorders of the coronary circulation, ischemia, reperfusion, low-intensity inflammation, dyslipidemia, pro-coagulation, stress, vascular and heart remodeling. Moebius strip like sequence of processing is represented as:

- Initialization of the composite symptoms, syndromes with geometry similar to Moebius strip, in particular that typical for pro-ischemic and anti-ischemic, reperfusion and anti-reperfusion, stress and anti-stress, hypoxic and antihypoxic, pro-oxidant and antioxidant, inflammatory and anti-inflammatory, dyslipidemic and anti-dyslipidemic, pro-arrhythmic and anti-arrhythmic, destructive and anti-destructive processes;
- Identification of Moebius strip chirality as a reflection of the predominance of pro-gradient or anti-gradient pathological changes;
- Clarification of Moebius strip constituents (symptoms, syndromes) as the object of research and its 3D representation. Temporal and spatial sequence of symptoms, syndromes reflects the pathogenetic and sanogenetic processes that allowed correcting treatment.

Our results of Moebius strip like pathology investigation were confirmed by determination of inflammation consumption of anti-inflammatory factors' syndrome as triggers of vascular pre-atherosclerotic remodeling in the patients with essential hypertension [6]. The purpose of this study was determination of the dependence between inflammation syndrome and vascular pre-atherosclerotic remodeling in the patients with essential hypertension (EH) [6]. The study included 36 patients (43-58 years old; men – 25, women - 12) with EH II with hypertensive heart [6]. We measured the interleukins – 1 beta, 6, 8, 10, C-reactive protein, factor of tumor necrosis alpha by ELISA. Holter electrocardiography monitoring and daily blood pressure monitoring was used [6]. Duplex Doppler-Echographic scanning of peripheral and common carotid arteries with the investigation of thickening of intima-media complex and diameter of blood vessels was performed in B-mode by device LOGIQ 400 (U.S.) [6]. Short list of research data is presented as: Med. – median; Q – Lower and Upper Quartiles. The level of CRP was increased in the patients with EH { Med. 0.06, Q (0,03-0,09)}, and patients with coronary heart disease in combination with EH {Med. 0.09, Q(0,02-0,13)}, in last somewhat more (Pmw = 0.035). The diagnostic criterions of consumption syndrome of pro-inflammatory and anti-inflammatory cytokines are characterized by the combination of cytokine imbalance and different levels of remodeling of common carotid and peripheral arteries and heart [6]. The first stage is characterized by the elevation of the level as pro-inflammatory (interleukin-1-beta - 50 pg/ml, interleukin-6 -

more than 5 pg/ml, interleukin-8 - more than 30 pg/ml, C-reactive protein - over 8.2 mg/l, tumor necrosis factor alpha - 50 pg / ml) so as anti-inflammatory cytokines (interleukin-10 - more than 70 pg/ml) with the minimal quantity of affected arteries, with following rise of pro-inflammatory and drop of anti-inflammatory cytokines (interleukin-10 - less than 3 pg/ml) and remodeling 3-4 arteries and more with thickening of intima-media complex from 1 to 1,29 mm [6]. Decreasing of pro-inflammatory cytokines and interleukin-10 determine maximal quantities (more than 4) of arteries, progression of thickening of intima-media complex at end stage [6]. Thus, we determined dependence between inflammation consumption syndrome and vascular pre-atherosclerotic remodeling in the patients with essential hypertension [6].

Our next investigation allowed to determine influence of inflammation consumption syndrome on myocardial electrical instability in the patients with chronic coronary heart disease and sinus node dysfunction [7]. The purpose of the study was determination the dependence between inflammation syndrome, geometry of depolarization of atria, ventricles and myocardial electrical instability in the patients with CHD [7]. The study included 36 patients (62,83± 1,49; 8,92 years old - M±SEM,SD; men - 27, women - 9 ) with chronic CHD, including 24 with stable angina pectoris (SAP), II-III functional class. Geometry of depolarization of atria, ventricles investigated in 26 patients with SAP II-IV functional with sick sinus node [7]. 36 healthy subjects (59,3±0,85; 11,22 years old; men - 24, women 12) consist control group [7]. We measured the interleukins - 1 beta, 6, 8, 10, C-reactive protein, factor of tumor necrosis alpha by ELISA [7]. Holter electrocardiography monitoring and daily blood pressure monitoring was used [7]. An increase of one or a combination of several pro-inflammatory (interleukin-1-beta - 50 pg/ml, interleukin-6 - more than 5 pg/ml, interleukin-8 - more than 30 pg/ml, C-reactive protein - over 8.2 mg/l, tumor necrosis factor alpha - 50 pg / ml) and decreased anti-inflammatory cytokines (interleukin-10 - less than 3 pg/ml) promotes atrial fibrillation and/ or group ventricular extrasystoles, or ventricular tachycardia in 21 patients from 36 (P<0,01 by criteria of sign) [7]. The Moebius like space orientation of depolarization processes were characterized by the change of supraventricular on ventricular pacemaker [7]. In 18 patients with sick sinus syndrome from 26 (P<0,01 by criteria of sign) were displayed Moebius strip like arrhythmias as a combination of supraventricular and ventricular extrasystoles; atrial fibrillation and group ventricular extrasystoles or ventricular tachycardia [7].

Ratio of pro-inflammatory and anti-inflammatory factors for acute coronary heart disease is mirror of Moebius strip like pathogenesis [8]. Our study included 27 patients (64,52±1,82;9,08 years old - y.o.) with ST segment elevation acute myocardial infarction (STEMI) and 25 (64,15±1,58; 8,23 y.o.) - with unstable angina pectoris (UAP) [8]. Research included investigation of cardiac biomarkers, interleukin-10 (IL-10), high sensitive C-reactive protein (hsCRP), auto-antibodies to chaperone 60 (anti-Hsp 60) [8]. Ratio on division of the level anti-Hsp 60 and hsCRP (anti-Hsp 60 / hsCRP) at patients with STEMI was

significantly lower {15,91±4,43; 23,04 conventional units (c.u.); (6,80-25,03); 5,73; (2,73-21,64), Pmw=0,0001} than in healthy 348,02±107,55; 340,12; (104,71-591,32); 275,33; (58,34-583,00). Index ratio of hsCRP to IL-10 {(hsCRP / IL-10) 1,50 ± 0,35; 1,83 c.u.; (0,78-2,22); 0,35; (0,07-2, 43) Pmw=0,001}; index of dividing the product of anti-Hsp 60 and hsCRP on IL-10 (anti-Hsp 60 \* hsCRP) / IL 10} at patients with STEMI was significantly higher 109,80 ± 31,48; 163,57 c.u.; (45,10-174,51); 32,17; (2,85-209,42), Pmw=0,001 compared with the control group [8]. In patients with UAP were dominant index ratio of IL-10 to the absolute phagocyte count (IL10/ absolute phagocyte count) (M±SEM;SD;95%CI: 2,82±0,27; 1,37; 2,26-3,39 c.u.), moderate this ratio at patients with uncomplicated STEMI (1,94±0,38;1,08;1,04-2,84 c.u.) and reduced - in complicated STEMI (1,59±0,26; 1,11; 1,05-2,13; PANOVA 1-2-3=0,006) [8]. A similar trend was typical for index ratio of hsCRP to the absolute phagocyte count (hsCRP / absolute phagocyte count) [8]. Thus, severe inflammatory and anti-inflammatory activities is characteristic feature for the patients with UAP [8]. Moderate inflammatory and autoimmune inflammatory activity with reduced anti-inflammatory potential was typical for patients with complicated STEMI [8].

An example of Moebius strip like atrial fibrillation treatment is presented lower. It is known that for the treatment of atrial fibrillation we use drugs that influence on sodium, potassium, calcium channels, beta-adrenergic receptors. Moebius strip like treatment effects may be in 2 variants:

1. The prevalence of antiarrhythmic effects - transformation atrial fibrillation to normal rhythm, normalizing quantity of heart rate;
2. Manifestations of pro-arrhythmic effects, transformation of atrial fibrillation to flutter.

#### 4. Conclusions

Our algorithm for determination of pathology mechanisms like a Moebius strip may be basis for making diagnosis and treatment decision. Modeling of this algorithm on Dragon language [12] help us to understand these processes better (scheme).

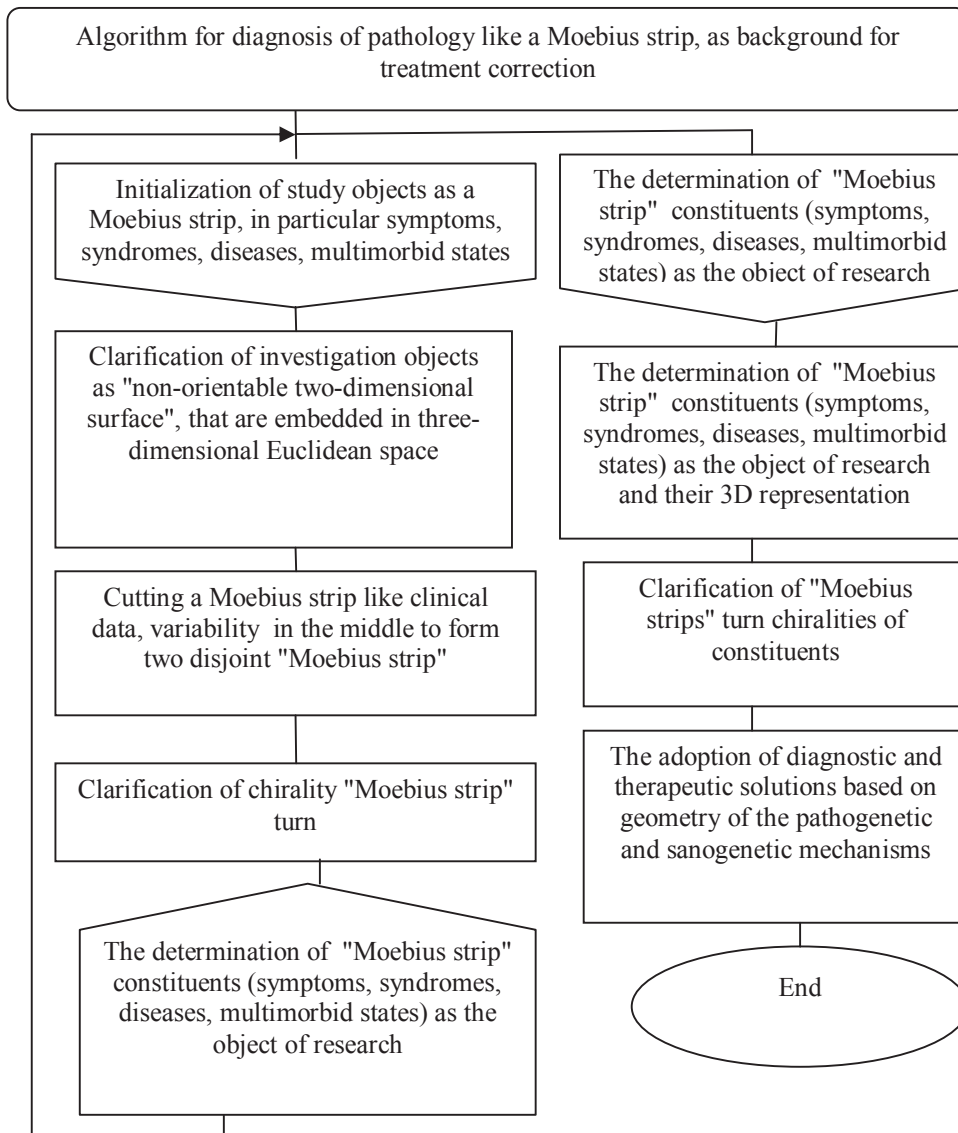
#### 5. References

- [1] S. K. Kulishov, Ye. A. Vorobjov, K.Ye. Vakulenko, A.G. Savchenko, T.I. Shevchenko, I.A. Latokha "Geometry of depolarization and repolarization processes in IHD patients of varying age with complete atrioventricular block or binodal disease as precondition to individualized treatment", J. Problems of aging and longevity, 2006, 15, № 4. - P. 332-338.
- [2] J. Zohar "Anxiety and depression: the Moebius strip", Medicographia, Servier publication, 2012, No 112, Vol. 34, No 3, P. 265-269.
- [3] Caballero Martínez "Does depression exist without anxiety?", Medicographia, Servier publication, 2012, No 112, Vol. 34, No 3, P. 318
- [4] S. Malyarov "Does depression exist without anxiety?", Medicographia, Servier publication, 2012, No 112, Vol. 34, No 3, P. 324.

- [5] J. Relvas "Does depression exist without anxiety?", *Medicographia*, Servier publication, 2012, No 112, Vol. 34, No 3, P. 326.
- [6] S.K. Kulishov, L.V. Solomatina. "Inflammation consumption of anti-inflammatory factors syndrome as trigger of vascular pre-atherosclerotic remodeling in the patients with essential hypertension", Final programme & Abstract Book of the European Conference. 2nd edition "Heart, Vessels & Diabetes", 3-5 November, Athens, Greece, 2011, P. 50.
- [7] S. Kulishov., K. Vakulenko, I. Latoha "Myocardial electrical instability as the derivative of inflammation consumption of anti-inflammatory factors syndrome, changes in geometry depolarization of atria, ventricles in the patients with coronary heart disease", Abstract book of "Rhythm 2011" Congress, 2011, P. 39.
- [8] S.K. Kulishov, N.P. Prikhodko "Ratio of pro-inflammatory and anti-inflammatory factors for acute coronary heart disease course", Supplement to Official Journal of the World Heart Federation "Global Heart" (World Congress of Cardiology Scientific Sessions, 2014, Incorporating the Annual Scientific Meeting of the Cardiac Society of Australia and New Zealand), 2014, March, Vol. 9, Issue 1S, e 282 (PW 109).
- [9] V.O. Bobrov., S.K. Kulishov "Adaptive ischemic and reperfusion syndromes in the patients with ischemic heart disease: mechanisms, diagnosis, substantiation of therapy", Poltava: Dyvosvit, 2004, 240 p.
- [10] S.K. Kulishov, O.M. Iakovenko "Solving clinical problems using system and anti-system comparison, graphic modeling"; *INNOVATIVE MEDICINE AND BIOLOGY*, Canadian International Monthly Reviewed Journal (CIJIMB), ISSN 1925-2188; 2011, No 3, P. 30-42.
- [11] Kulishov S.K., Vakulenko K.Ye., O.M. Iakovenko "Peculiarities of cardiac remodeling, cytokines change, electrical myocardial instability in patients with chronic ischemic heart disease and arterial hypertension as predispose to making treatment decision" // Supplement to Official Journal of the World Heart Federation "Global Heart" (World Congress of Cardiology Scientific Sessions, 2014, Incorporating the Annual Scientific Meeting of the Cardiac Society of Australia and New Zealand), 2014, March, Vol. 9, Issue 1S, e 169 (PT 023).
- [12] Parondzhanov V.D. "Learn to write, read and understand algorithms. Algorithms for the correct thinking. Algorithmic Basics", M.: DMK Press, 2012. – 520 p.

**The scheme**

Algorithm for diagnosis of pathology like a Moebius strip, as background for treatment correction





# Identifying severe trauma in the ED with a RIMER methodology

Guilan Kong<sup>1,\*</sup>, Jian-Bo Yang<sup>2</sup>, Dong-Ling Xu<sup>2</sup>, Richard Body<sup>3</sup>, Xiaofeng Yin<sup>4</sup>, Tianbing Wang<sup>4</sup>,  
Baoguo Jiang<sup>4</sup>, Jing Wang<sup>1</sup>, Yonghua Hu<sup>1</sup>

<sup>1,\*</sup>: Medical Informatics Center, Peking University, Beijing, China

<sup>2</sup>: Decision and Cognitive Sciences Research Centre, The University of Manchester, Manchester, UK

<sup>3</sup>: Emergency Department, Manchester Royal Infirmary, Oxford Road, Manchester, UK

<sup>4</sup>: Traffic Medical Center, Peking University, Beijing, China

**Abstract** - A generic belief rule-based inference methodology using the evidential reasoning approach (RIMER) was employed in this study to develop a clinical decision support system (CDSS) intended to aid physicians to identify severe trauma in the emergency department (ED). The RIMER methodology extends traditional IF-THEN rules to belief rules with belief degrees embedded in all possible consequents. The RIMER system has advantages in representing uncertain relationships between rule antecedents and consequents. The five vital signs, including body temperature, respiratory rate, systolic blood pressure, pulse rate, and the level of consciousness, were used as antecedent factors in the study to assess trauma severity. Five-fold cross-validation method was employed to train and validate the system. This study contributes to the trauma CDSS research area with an innovative RIMER system, which provides a pragmatic way to make full advantages of expert knowledge and historical data in helping ED physicians identify severe trauma.

**Keywords:** Trauma, Clinical decision support system, RIMER, Uncertainty

## 1 Introduction

Trauma has become one of the leading causes of mortality and disability worldwide. Trauma accounts for 16% of the global burden of disease, and 16,000 people die from injury daily. Physicians in the emergency department (ED) need to provide a rapid initial assessment of illness severity for trauma patients just after their arrival to hospital, so as to make appropriate decisions about triage and treatment of trauma. Particularly, for better patient outcomes and the optimal utilization of hospital resources, the trauma patients with high probability of major adverse events, including in-hospital death and intensive care unit (ICU) admission, need to be identified first. In the literature, research [1] shows that vital signs are fundamental factors for trauma severity assessment and the important predictors of in-hospital death and ICU admission. The nature of existent trauma severity assessment tools is to assign a severity score to patients based on observations or the measured values of vital signs.

However, the relationship between vital signs and severe trauma cannot be represented explicitly by the existing scoring tools. To explore the relationship between vital signs and severe trauma, which results either in in-hospital death or in ICU admission, logistic regression (LR) [2, 3] and artificial neural network (ANN) [4] were employed by researchers to construct prediction models. Although both LR and ANN do not require the concrete knowledge of the relationship between independent factors and dependent outcomes, both methods need large enough sample data to learn prediction models, and both models are black box for physicians.

In this study, we proposed to employ a generic belief rule-based inference methodology using the evidential reasoning approach (RIMER) [5] to develop a clinical decision support system (CDSS) to aid physicians to identify severe trauma in the ED.

## 2 Materials and methodology

### 2.1 Data

A sample of 1,190 trauma patients directly sent to the ED at Kailuan Hospital, North China, between 2008 and 2009 was employed for system development and validation. The primary outcome was severe trauma composing of both in-hospital death and the ICU admission.

### 2.2 The RIMER methodology

In the RIMER methodology, traditional IF-THEN rules are extended to belief rules by embedding belief degrees in all possible consequents of a rule. Meanwhile, other knowledge representation parameters including rule weights and antecedent attribute weights are embedded in the belief rules. Inference with belief rule base (BRB) in a RIMER system is implemented using the evidential reasoning (ER) approach. The RIMER system has the advantages of using belief rules to represent clinical domain knowledge under uncertainty and inference with uncertain clinical data using the ER approach. Moreover, the knowledge representation parameters, including rule weights, antecedent attribute

weights, and consequent belief degrees in the BRB, can be fine-tuned or trained using accumulated historical data [6]. The RIMER methodology has been employed to stratify patients with cardiac chest pain [7]. A brief introduction of RIMER is as follows. A belief rule can be described as  $R_k$  :

$$\text{If } A_1^k \wedge A_2^k \wedge \dots \wedge A_{T_k}^k, \text{ Then } \{(D_1, \beta_{1k}), (D_2, \beta_{2k}), \dots, (D_N, \beta_{Nk})\} \\ \left( \beta_{jk} \geq 0, \sum_{j=1}^N \beta_{jk} \leq 1 \right), \text{ with a rule weight } \theta_k \text{ and attribute weights } (1) \\ \delta_1, \delta_2, \dots, \delta_{T_k} \quad k \in \{1, \dots, L\}$$

where  $A_i^k (i=1, \dots, T_k)$  is the referential category or grade of the  $i$  th antecedent attribute used in the  $k$  th rule;  $\beta_{jk} (j=1, \dots, N; k=1, \dots, L)$  the belief degree assigned to consequent  $D_j$  and it is originally given by experts;  $\delta_i (i=1, \dots, T_k)$  the antecedent attribute weight representing the relative importance of the  $i$  th attribute;  $\theta_k$  the rule weight representing the relative importance of the  $k$  th rule.  $L$  is the number of all belief rules in the rule base.  $T_k$  is the number of all antecedent attributes used in the  $k$  th belief rule.  $N$  is the number of all possible consequents in the BRB. Traditional IF-THEN rules can be represented as a special case of BRB with only one consequent and the consequent belief degree is 100%. The inference with the BRB in the RIMER system is implemented using the ER algorithm [8, 9]. Firstly, transform numerical values or subjective judgments associated with clinical variables to belief distributions on corresponding reference categories used in the BRB. Secondly, calculate the activation weight of each belief rule in the BRB on the basis of degrees of belief in various reference categories in the transformed clinical inputs and the weight associated with each rule. Thirdly, use the ER algorithm to aggregate all the activated belief rules in the BRB, thus generating a distributed trauma severity assessment with combined belief degrees distributed on all possible consequents. The analytic ER algorithm [10] is as follows.

$$\beta_j = \frac{\mu \times \left[ \prod_{k=1}^L \left( a_k \beta_{jk} + 1 - a_k \sum_{f=1}^N \beta_{fk} \right) - \prod_{k=1}^L \left( 1 - a_k \sum_{f=1}^N \beta_{fk} \right) \right]}{1 - \mu \times \left[ \prod_{k=1}^L (1 - a_k) \right]}, \quad j=1, \dots, N \quad (2)$$

where

$$\mu = \left[ \sum_{f=1}^N \prod_{k=1}^L \left( a_k \beta_{fk} + 1 - a_k \sum_{j=1}^N \beta_{jk} \right) - (N-1) \times \prod_{k=1}^L \left( 1 - a_k \sum_{j=1}^N \beta_{jk} \right) \right]^{-1}, \\ \omega_k \text{ is the rule activation weight calculated in the second step, and } \beta_j (j=1, \dots, N) \text{ is the final combined belief degrees associated with corresponding consequents } D_j (j=1, \dots, N).$$

Initial belief rules in the study were provided by domain experts, and the five vital signs including body temperature,

respiratory rate, systolic blood pressure, pulse rate, and the level of consciousness were used as antecedent factors in the rule base. Possible consequents of the rule base are “severe trauma” and “non-severe trauma”. Here severe trauma is defined by using in-hospital death and ICU admission as criteria. The referential grades of the five vital signs are “normal (N)” and “abnormal (AN)”. One example belief rule is “IF body temperature is AN, respiratory rate is AN, systolic blood pressure is AN, pulse rate is AN, and the level of consciousness is AN, THEN {(severe trauma, 70%), (non-severe trauma, 30%)}”, where {(severe trauma, 70%), (non-severe trauma, 30%)} means that if one patient’s clinical data satisfy the compact antecedent of the rule, the patient will have in-hospital death or ICU admission with 70% certainty, and there is 30% certainty that the patient will have neither of the two adverse events. Totally, there were 32 belief rules in the BRB as shown in Table 1, where  $A^1$  represents body temperature,  $A^2$  respiration rate,  $A^3$  pulse rate,  $A^4$  systolic blood pressure, and  $A^5$  represents the level of consciousness.

Subjective judgments of consciousness and numerical measurements of other four vital signs can be transformed to ‘AN’ or ‘N’ to some degree. Inference with the transformed inputs in the BRB was implemented using the ER algorithm [5]. A final combined severity assessment with belief degrees distributed on “severe trauma” and “non-severe trauma” was generated for each case that was inputted into the system.

As those initial rules were provided by domain experts, and they may not represent the real situation with 100% accuracy, the knowledge representation parameters need to be fine-tuned or trained using accumulated historical data. We employed five-fold cross-validation method to train and validate the RIMER system. The dataset was split into five folds with equal cases (containing not only patients with negative outcomes but also those with positive outcomes), and the BRB was trained and tested five times using any four folds as training set and the remaining fold as test set. The objective of the BRB training was to find out an optimal set of knowledge representation parameters for the BRB through minimizing the discrepancies between system generated outputs and the observed outcomes for all cases in the training data set.

Finally, considering the observed outcome as gold standard, the final inferred belief degrees distributed on severe trauma generated by the RIMER system in each test fold were used to plot the receiver operating characteristic (ROC) curve, and the area under the curve (AUC) was used as a measure to compare the prediction performance of the RIMER system before and after BRB training.

**Table 1.** Initial belief rules provided by experts

No.	Antecedent					Consequent
	A <sup>1</sup>	A <sup>2</sup>	A <sup>3</sup>	A <sup>4</sup>	A <sup>5</sup>	Trauma severity
1	AN	AN	AN	AN	AN	{(severe, 70%), (non-severe, 30%)}
2	AN	AN	AN	AN	N	{(severe, 60%), (non-severe, 40%)}
3	AN	AN	AN	N	AN	{(severe, 60%), (non-severe, 40%)}
4	AN	AN	AN	N	N	{(severe, 50%), (non-severe, 50%)}
5	AN	AN	N	AN	AN	{(severe, 60%), (non-severe, 40%)}
6	AN	AN	N	AN	N	{(severe, 50%), (non-severe, 50%)}
7	AN	AN	N	N	AN	{(severe, 50%), (non-severe, 50%)}
8	AN	AN	N	N	N	{(severe, 30%), (non-severe,70% )}
9	AN	N	AN	AN	AN	{(severe, 60%), (non-severe, 40%)}
10	AN	N	AN	AN	N	{(severe, 50%), (non-severe, 50%)}
11	AN	N	AN	N	AN	{(severe, 50%), (non-severe, 50%)}
12	AN	N	AN	N	N	{(severe, 30%), (non-severe,70% )}
13	AN	N	N	AN	AN	{(severe, 50%), (non-severe, 50%)}
14	AN	N	N	AN	N	{(severe, 30%), (non-severe,70% )}
15	AN	N	N	N	AN	{(severe, 30%), (non-severe, 70% )}
16	AN	N	N	N	N	{(severe, 10% ), (non-severe, 90%)}
17	N	AN	AN	AN	AN	{(severe, 60%), (non-severe, 40%)}
18	N	AN	AN	AN	N	{(severe, 50%), (non-severe, 50%)}
19	N	AN	AN	N	AN	{(severe, 50%), (non-severe, 50%)}
20	N	AN	AN	N	N	{(severe, 30%), (non-severe, 70% )}
21	N	AN	N	AN	AN	{(severe, 50%), (non-severe, 50%)}
22	N	AN	N	AN	N	{(severe, 30%), (non-severe, 70% )}
23	N	AN	N	N	AN	{(severe, 30%), (non-severe, 70% )}
24	N	AN	N	N	N	{(severe, 10%), (non-severe, 90%)}
25	N	N	AN	AN	AN	{(severe, 50%), (non-severe, 50%)}
26	N	N	AN	AN	N	{(severe, 30%), (non-severe, 70% )}
27	N	N	AN	N	AN	{(severe, 30%), (non-severe,70% )}
28	N	N	AN	N	N	{(severe, 10%), (non-severe, 90%)}
29	N	N	N	AN	AN	{(severe, 30%), (non-severe, 70%)}
30	N	N	N	AN	N	{(severe, 10%), (non-severe, 90%)}
31	N	N	N	N	AN	{(severe, 10%), (non-severe, 90%)}
32	N	N	N	N	N	{(severe, 5%), (non-severe, 95%)}

**Table 2.** AUC of the RIMER system before and after training

Folds	1 <sup>st</sup> fold	2 <sup>nd</sup> fold	3 <sup>rd</sup> fold	4 <sup>th</sup> fold	5 <sup>th</sup> fold	Average
<b>Initial</b>	0.821	0.993	0.912	0.973	0.981	0.936
	95%CI:	95%CI:	95%CI:	95%CI:	95%CI:	
	0.767-0.868	0.972-0.999	0.868-0.944	0.944-0.990	0.955-0.994	
<b>After training</b>	0.850	0.995	0.940	0.988	0.985	0.952
	95%CI:	95%CI:	95%CI:	95%CI:	95%CI:	
	0.798-0.893	0.975-1.000	0.902-0.967	0.964-0.998	0.961-0.996	

CI: Confidence Interval

### 3 Results

Using actual observed outcome for included patients as benchmark data, the AUC which represents prediction performance of the RIMER system before and after system training for each fold is shown in Table 2.

### 4 Conclusions

Based on the study results, it can be found that the initial BRB provided by experts provided a good prediction performance, and the system performance of the RIMER CDSS was robust and can be improved to some degree after being fine-tuned with historical trauma data.

To conclude, the RIMER methodology provides an innovative way for CDSS researchers to make full advantages of expert knowledge and historical patient data to develop a decision support tool to identify severe trauma. The RIMER system has a strong potential to aid ED physicians to better triage trauma, optimally utilize hospital resources, and achieve better patient outcomes.

### 5 Acknowledgements

This study was supported by a humanities and social sciences project funded by the Ministry of Education of China under Grant no.: 13YJC630066. The study was also supported by a grant from the National Natural Science Foundation of China (NSFC) under Grant no.:81301296.

### 6 References

- [1] T. Smith, *et al.*, "Accuracy of an expanded early warning score for patients in general and trauma surgery wards," *British Journal of Surgery*, vol. 99, pp. 192-197, 2012.
- [2] S. Alghnam, *et al.*, "Predicting in-hospital death among patients injured in traffic crashes in Saudi Arabia," *Injury*, vol. 45, pp. 1693-1699, 2014.
- [3] W. K. Utomo, *et al.*, "Predictors of in-hospital mortality and 6-month functional outcomes in older adults after moderate to severe traumatic brain injury," *Injury*, vol. 40, pp. 973-977, 2009.
- [4] S. M. DiRusso, *et al.*, "Development of a model for prediction of survival in pediatric trauma patients: comparison of artificial neural networks and logistic regression," *Journal of Pediatric Surgery*, vol. 37, pp. 1098-1104, 2002.
- [5] J. B. Yang, *et al.*, "Belief rule-base inference methodology using the evidential reasoning approach - RIMER," *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans*, vol. 36, pp. 266-285, Mar 2006.

[6] J. B. Yang, *et al.*, "Optimization models for training belief-rule-based systems," *IEEE Transactions on Systems, Man and Cybernetics, Part A-Systems and Humans*, vol. 37, pp. 569-585, 2007.

[7] G. L. Kong, *et al.*, "A belief rule-based decision support system for clinical risk assessment of cardiac chest pain," *European Journal of Operational Research*, vol. 219, pp. 564-573, 2012.

[8] J. B. Yang and M. G. Singh, "An evidential reasoning approach for multiple-attribute decision making with uncertainty," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 24, pp. 1-18, 1994.

[9] J. B. Yang and D. L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans*, vol. 32, pp. 289-304, May 2002.

[10] Y. M. Wang, *et al.*, "Environmental impact assessment using the evidential reasoning approach," *European Journal of Operational Research*, vol. 174, pp. 1885-1913, 2006.

## **SESSION**

# **ELDERLY CARE: SYSTEMS, METHODS, AND NOVEL PRACTICES + ASSISTIVE TECHNOLOGIES**

**Chair(s)**

**TBA**



# A Smart Shoe to Prevent and Manage Diabetic Foot Diseases

<sup>1</sup>Olawale David Jegede, <sup>1</sup>Ken Ferens, <sup>2</sup>Bruce Griffith, <sup>3</sup>Blake Podaima

<sup>1</sup>Dept. of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada

<sup>2</sup>Avriel International Inc., Winnipeg, MB, Canada

<sup>3</sup>Internet Innovations Center, Winnipeg, MB, Canada

Ken.Ferens@umanitoba.ca

*Abstract—Diabetes is a global disease which is growing rapidly. The Canadian Diabetes Association put the estimate of Canadians living with this disease at more than 9 million (which is about a quarter of the nations' population) while the World Health Organisation puts the global prevalence as one out of twelve people. If not managed carefully, the disease can lead to further complications resulting in death. There has been several research efforts aimed at managing diabetes in order to prevent possible complications; such complications could be to any part of the body. In the past we proposed a solution to detect, prevent and manage complications that may result in the foot; the solution is a computer engineering approach that uses wireless technology and intelligent interpretation of sensed data to detect, prevent and manage possible chronic diabetic foot diseases. This work is an improvement on the proposed computer engineering approach.*

**Keywords—** Wireless Networks, Local Area Network, Wide Area Network, Sensors, Sensory Neuropathy, Diabetes Mellitus, Operating System, Application, Readings, Data.

## 1. Introduction

Diabetes is a disease in which the human body is unable to handle blood glucose (sugar). The pancreas is an organ situated at the back of the stomach; it releases a hormone called insulin which is used by the body when the blood glucose level is deemed to be considerably higher than needed in the blood stream. This insulin is able to convert the excess glucose into energy for use by the body or sometimes store the energy in the body tissues, such as body fat and muscles. When this metabolic process breaks down and the pancreas is

producing little or no insulin, it results in the body having too much blood glucose needed for the body to stay healthy. This disease is what is known as diabetes.

Currently, the Canadian Diabetes Association estimates that more than nine million people in Canada have diabetes or pre-diabetes; this accounts for about a quarter of the total Canadian population, meaning that one in four Canadians have the disease. Thus, for every Canadian, diabetes affects the person or someone that the person knows. The International Diabetic Federation (IDF) estimated that in 2014, the number of people with diabetes was 387 million with a prevalence rate of 8.3%; this figure is projected to increase to about 592 million [1] by the year 2035. The IDF also reported that about 1 in 2 people globally do not know that they are diabetic and that there are more people living with diabetes within the age 40 – 59 years. The costs of treating and managing diabetes in North America and the Caribbean accounts for about half of the global expenditure; in 2013 this was reported to be about \$263 billion. The lifestyle of the residents of these two regions is responsible for the reported prevalence with respect to other parts of the world. The Canadian Diabetes Association (CDA) estimates that currently, over nine million Canadians live with diabetes or pre-diabetes [2], meaning that one in every four Canadians is affected. In the year 2010, the CDA estimated that about 94,000 people live with diabetes in Manitoba, Canada with a prevalence rate of 7.6 %; the figures are projected to increase to 139,000 and 10.1 % respectively by the year 2020. The estimated cost of treatment for Manitoba patients was also reported to be \$498 million in 2010 and projected to increase to about \$639 million by 2020.

One of the ways to reduce the costs of the disease is to find a way to prevent possible associated complications.

Our work focuses on preventing the possible foot complications. *Sensory neuropathy* is a common side effect suffered by a majority of diabetic patients. It is a type of peripheral neuropathy that results in damage to the human body nerve; this causes loss of sensation in the body and sets the body up for further complications. The part of the body that suffers more from this anomaly is the foot; the foot is unable to feel injury or change in environmental parameters such as temperature, humidity and pressure. Armstrong et al [3] reported 15% of the patients end up with *foot ulcers* while 85% end up needing *surgical amputation* as a result of the loss of sensation. Also, about 100,000 limbs are lost each year in the United States to complications from loss of sensation [3]. An injury to the foot of a diabetic doubles the possibility of death within when compared to a non-diabetic. A key indicator of injury on any part of the foot is inflammation in the affected part. Without having to view the foot, an injured part can be distinguished from non-injured part by taking measurement of three parameters across the foot; these parameters are *temperature*, *mechanical (plantar) pressure*, and *humidity*. We have developed a smart shoe that monitors these parameters across the foot and alerts the user about any significant change in the measurement of any of the parameters across the foot. In this way the patient can better self-monitor the condition of the foot and ensure prevention of any complication.

The remaining sections of this paper are organized as follows. Section 2 discusses related works done in this research area while Section 3 focuses more on the computer engineering approach. In our previous work, we have used sensors with certain limitations, but this new work utilizes sensors without those limitations; Section 3 sheds more light on this. Section 4, discusses the results obtained while Section 5 discusses possible limitations. Section 6 concludes the paper and gives possible future work/direction.

## 2. Related Work

There are previous works that have been done in this area to detect signs of injury in diabetic patients thus preventing complications. Zequera et al [4] and Kanade [5] utilized the plantar (mechanical) pressure measurements to monitor and diagnose the foot of a diabetic patient. Frykberg et al [6], Armstrong [7] and Lavery et al [8] made use of the temperature measurement for their diagnosis. They used the parameter to determine part of the foot that may have inflammation.

Frykberg et al [6] developed a device named Tempstat™, this is a self-monitoring, easy-to-use device which patients use to examine the plantar surface of their foot to detect any sign of injury. Armstrong [7] developed a self-monitoring home-based dermal thermometry to also detect injury; he was able to demonstrate that using the device reduced foot ulceration especially in high-risk diabetic patients. Lavery et al [8] also showed that with the use of a self-monitoring at-home infrared temperature device, a diabetic patient can effectively prevent foot complications. Other work in literature pointed to the use of the previously discussed parameters (plantar pressure and temperature) to monitor diabetic's foot condition. In this work, we have included a third parameter that can also be used to monitor foot conditions; this parameter is *humidity*. Humidity is the amount of water vapor available in the air; relative humidity is expressed as a percentage of the amount needed for saturation at the same temperature. Relative humidity is directly related to temperature; thus it affects how the human body feels with respect to the environment. When the relative humidity is high, we feel much hotter than the actual temperature; this is also true vice versa. According to the Center for Disease Control and Prevention [9], diabetes hampers the ability of the body to cope with high heat and humidity. Thus, our method uses the three parameters to measure the foot condition, thus increasing accuracy and sensitivity of the system.

## 3. Computer Engineering Approach

In our previous work [10] we proposed a computer engineering approach that is able to take the measurement of the combination of the three parameters - *temperature*, *pressure* and *humidity* - across the foot in order to prevent foot ulceration. The work is centered on being able to detect foot inflammation which is a sign of danger. Based on the SurroSense Rx System [3], we identified 8 parts (hot spots) of a typical human foot that can be monitored; the parts are such that the distribution across the foot can be said to be uniform. For each hot spot we measure the three parameters. The ideal is to have a relatively similar measurement of each of the parameter across the hot spots. A deviation from the ideal portends danger on the foot. The 8 hot spots are situated as follows: one hotspot at the *big toe*, three hot spots at the *third toe*, three hot spots *across the ball of the foot*, two hot spots in the *arch*, and one hot spot at the *heel*. A pictorial representation of this distribution is given in [10]. Our system model is described in the next section.



### 3.1. System Model

Our system model is shown in Fig. 1. The smart shoe contains sensors that obtain readings (data) for each of the three parameters, and at each of the hot spots. The data is then sent through an attached Bluetooth radio to a smart phone that displays the data on an app for the purpose of self-monitoring. The app then sounds an alarm as soon as considerably large difference is received for any of the parameters at two or more hot spots. The data is then sent over the wireless network (LAN/WAN) to a central server where further processing and interpretation is done; the goal is to make this information accessible to health officials for the purpose of diagnosis and monitoring.

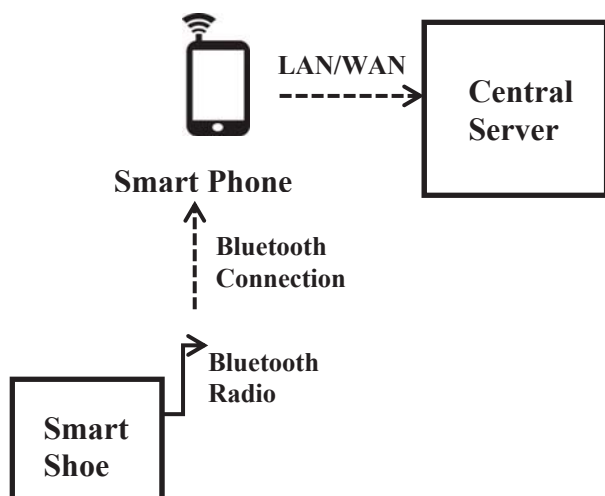


Fig. 1 System model [10].

• **Smart Shoe:** The word ‘smart’ in the smart shoe comes as result of the fact that the ‘sensing’ electronics is contained within the shoe. The insole for the shoe contains the sensors (temperature, humidity and pressure); that were carefully embedded in. The main difference between our previous work [10] and this new work lies in the sensors/components used in the smart shoe; the difference is given by Table 1.

Table 1 Component differences.

Parameter	Previous Work [10]	New work
Temperature	DHT22	HIH6130
Humidity	DHT22	HIH6130
Plantar pressure	Strain Gauge	FlexiForce

The DHT22 is able to measure both the temperature and humidity but it occupies a large space and can only get data once every 2 seconds [11]. Also it is not I2C compatible and as such is not addressable. On the other hand, the HIH6130 is smaller and can get data in less than a second; it can also measure a wider range of temperature (-25 to 85C) and humidity (0-100%) measurement. More importantly, it allows I2C bus communication. The ATmega328 Arduino microcontroller was used to control and coordinate the computing processes. The sensor specification, pin layout and connection can be found in [12]. For our application, the pins 5, 6 & 7 of the hih6130 was not needed. A power supply of 5V dc and GND were supplied by the Arduino board. The pins 3 and 4 of each of the 8 sensors were connected to the SCL (clock) and SDA (data) lines respectively. The SCL is the clock line bus that is used to synchronize communication between the devices; it is controlled by a master controller. The SDA is the data transfer bus on which the devices send their readings. The Arduino mega 2560 provides a Two Wire Interface (TWI) or I2C interface via the digital pins 20 for SDA and 21 for the SCL. Fig. 2 shows a typical I2C connection for data transfer from *n* devices to a controller.

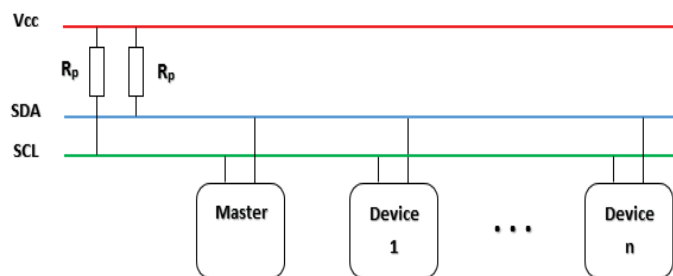


Fig. 2 I2C block diagram.

The symbol *R<sub>p</sub>* in Fig. 2 represents a pull up resistor which ensures that the input pin reads a high state when none of the devices is active; *V<sub>cc</sub>* represents voltage supply. The Master device controls the whole process of data transfer by querying the devices one at a time for the sensed readings real time. The Master device for our application is the embedded ATmega2560 on the Arduino board. Each of the other devices apart from the master is regarded as a slave. In our application we have 8 devices (hih6130 sensors) as the slaves to be queried. The ATmega2560 is used to hardcode an I2C address for each of the sensors; this I2C address is in the form of integer representation from 1 to 8. We have adapted the work done by Casagrande [13] for our application and we have integrated the code for obtaining readings from the FlexiForce.

The Strain gauge and FlexiForce sensors are able to measure applied plantar pressure using resistance-based technologies. However, the FlexiForce provides more advantages to the strain gauge in the following respect: the FlexiForce is more sensitive, smaller in size, requires simpler electronics and is easier to integrate than the strain gauge [14]. Thus, we have chosen the FlexiForce [15] for this work. The resistance of the sensing element of the FlexiForce changes in inverse proportion to the force being applied on the active sensing area of the sensor [16]. We have made use of the A201 model that gives an analog output. We have incorporated the FlexiForce into a force-to-voltage circuit being driven by a 5V supply. The analog output is supplied to the analog input pins of the Arduino mega 2560 board. Since we have 8 hotspots for our applications, we have made use of 8 FlexiForce sensors and supplied the output of each to one of the analog input pins A0 – A7 of the Arduino board. The Arduino board has an embedded analog-to-digital converter that is used to change the analog input to a digital (voltage) output from the board. In order to increase the sensitivity of the sensor, we have made use of a  $1M\Omega$  resistor as the reference voltage. The connection guide is given in [17]. The same connection applies to the remaining 7 FlexiForce sensors except that each one connects to different analog input port (A1 – A7) on the Arduino board.

• **PCB Fabrication:** We have dimensioned each hot spot by taking into account the average adult foot size and the size of the sensors available. The dimensioned PCB is 0.02m by 0.02m in size and there are 8 of them, one per hotspot. This accommodates both the hih6130 sensor and the FlexiForce sensor. The hih6130 is soldered on one side of the PCB while the pressure sensor is attached to the other side. We have embedded the PCB into the shoe such that hih6130 is facing up while the pressure sensor faces down; by so doing the hih6130 is able to take the accurate and actual readings of the respective part of the foot while the pressure sensor under the PCB is able to accurately measure the plantar force. An example of our fabricated PCB is given by Fig. 3.



Fig. 3 PCB with mounted hih6130 and FlexiForce.

Fig. 3 (left side) shows the PCB with a mounted hih6130 sensors having a paper label 7; this label indicates the number that has been hard-coded into the sensors for the purpose of communication on the I2C bus. Each of the 8 hih6130 sensors is hardcoded with a number that fall between the range 1 to 8, and no two sensors can have the same number. Thus the number 7 indicates that this is the 7th sensor (out of the 8 sensors) that is queried by the processor for its readings. Please note that this label is removed when the PCB is to be embedded into the sole of the shoe. The label was just to identify each sensor and its position on the bus. Fig. 3 (right side side) shows the PCB with a FlexiForce that has been glued to the other side of the PCB. There was no reason to label the pressure sensors since their output is analog and each one is connected to analog ports on the Arduino. Thus, the 0.02m<sup>2</sup> sized PCB is big enough to have the hih6130 and FlexiForce sensor mounted on it. We did this for the 8 PCBs and connected the sensors to the Arduino board according to the descriptions in the sensors data sheet.

• **Insole & Embedded PCBs with Sensors:** The position of each hot spot was decided with respect to the description in Section 2. These hotspots were in turn cut out and the fabricated PCBs with sensors were embedded at the hotspots. The end product is given by Fig. 4 which shows Flexi Force part of the PCB while the hih6130 part of the PCB is facing the inside of the insole. The wires coming from the PCBs are then connected to the Arduino board based on the descriptions in the data sheets. The hih6130 sensors are connected in parallel because of the I2C capability.



Fig. 4 Insole with embedded sensors.

• **HC – 06 Serial Bluetooth Brick:** In this work, we have made use of the Bluetooth wireless technology for data transmission because majority of smart phones today have Bluetooth capability. Bluetooth capability enables wireless connectivity between devices. A Bluetooth radio was attached to the smart shoe in order to allow the sensor readings to be transmitted to a mobile phone where

the information can be viewed and interpreted for self-monitoring use by the patient. Smart phones generally support operating systems including iOS, Android, Blackberry and Windows. Each of these operating systems today supports Bluetooth low energy. Bluetooth is able to support a relatively shorter propagation range (1-100m) unlike other wireless communication standards. The HC-06 bluetooth module was used for this work because of its ease of use and wide range of operating system versions that it supports. It enables the sensor readings to be sent to the smart phone.

- **Smart Phone App:** We developed a self-monitoring android app using the Android Development Tool (ADT) on an Eclipse Integrated Development Environment (IDE). The app was developed by modifying an existing app that connects through Bluetooth to an Arduino and displays temperature readings [18]. The developed app was installed on an android-based smart phone (Samsung Galaxy S3 Mini). The app was programmed to use the bluetooth capability of the smart phone to receive data (sensor readings) from the bluetooth enabled-smart shoe. The first step is to launch the app on the phone and then connect it to the bluetooth protocol on the phone; in the event that the bluetooth is not turned on, the app automatically prompts the user to turn bluetooth on. The user now has the option to pair the app with any other bluetooth device that is within the phone's bluetooth range. A pictorial view of the app is given in Fig. 5. In this case the bluetooth-enabled app is not yet paired with any other bluetooth device; thus showing as "not connected". In our case, we paired the app with the HC-06 Serial Bluetooth brick (bluetooth module attached to the smart shoe) - specifically programmed with the name "itead 00:13:12:31:20:46". Once the pairing is successful, the item "not connected" on the app changes to "connected" and the readings from each hotspots of the shoe is displayed on corresponding positions on the app.

The readings from a particular hotspot position are displayed on corresponding position on the app. The data being displayed on the app is stored in a file on the phone.

- **Central Server:** A server is software that runs on a computer (hardware) to serve the requests made by other programs running on a set of other computing devices – in our case smart phones - on a computing network (LAN/WAN). The server is proposed to receive data (sensor readings being displayed on the app) from the phone for the purpose of data storage and interpretation by health personnel.



Fig. 5 GUI of the smart phone app.

#### 4. Experimental Results

Table 2 shows the readings obtained for each of the three parameters at each of the 8 hotspots at a given time. The data obtained per parameter are similar with slight difference which can be attributed to the position of the hotspots that has already been embedded in the shoe. The temperature reading seems more distributed across the hotspots and represents the room temperature measurement. The humidity reading for the hotspots towards the lower part of the foot (heel and arch) is higher – this can be attributed to the opening of the shoe (through which humans insert their foot) at this position. The pressure readings also varies somewhat uniformly across the hotspot; it is important to note that the pressure readings were taken when there was no human foot inserted inside the shoe – that is responsible for the lower pressure value. The value would normally range between 4.0 and 5.0 when a patient's foot is in the shoe and the patient is walking and/or exerting pressure.

Table 2 Readings for the 8 Hotspots at a particular time.

Hotspot #	Parameter		
	Temperature (° C)	Humidity (%)	Pressure
1	23.25	46.46	0.71
2	22.34	51.40	0.11
3	20.85	48.49	0.43
4	21.01	50.01	0.66
5	21.02	55.68	0.97
6	20.98	56.07	1.21
7	21.62	55.68	1.20
8	23.85	54.05	0.66

Table 3 shows the readings obtained for each of the three parameters for the 8<sup>th</sup> hotspot over a period of time, specifically about 5 minutes. The shoe was powered on for about 5 minutes with each hotspot readings for the three parameters obtained every 0.5 seconds. The readings were taken at the heel position on the sole, which also represents the 8<sup>th</sup> hotspot. During this experiment, a soldering iron was heated up for about 10 minutes and then placed at the 8<sup>th</sup> hotspot position for a period of about 5 minutes. The readings show a gradual increase in the temperature measurement from an initial (current room temperature) value of 23.85 °C to a high value of 35.89 °C at which time the soldering iron then begins to cool off and the temperature begins to drop as expected. The humidity readings also follow the same pattern proportionately as the temperature except for a few disparate readings. As expected the pressure readings varied insignificantly and could be said to stay the same on the average; this was due to absence of any pressure on the sole.

Table 3 Hotspot 8's (Heel) readings over 5 min.

Reading #	Parameters		
	Temperature (°C)	Humidity (%)	Pressure
1	23.85	54.05	0.66
2	24.73	45.68	0.66
3	25.92	40.08	0.67
4	26.90	38.01	0.66
5	27.98	40.01	0.65
6	29.39	42.78	0.67
7	30.43	44.01	0.65
8	31.15	49.59	0.66
9	31.48	48.57	0.67
10	32.05	53.47	0.65
11	34.12	56.64	0.66
12	34.80	60.62	0.68
13	35.89	62.61	0.65
14	31.98	59.44	0.66
15	30.40	60.08	0.65
16	29.59	59.71	0.65
17	27.19	61.20	0.65
18	27.07	61.05	0.65
19	27.04	61.00	0.64
20	27.00	60.95	0.65

We have plotted each of the parameters' readings over the time period for which the measurements were taken; this is given by Fig. 6 which shows that the temperature and pressure readings follow the same pattern over time while the humidity readings remained constant.

## 5. Potential Limitations

In our previous work [10], we highlighted some potential limitations in the usage or development of this product. Patients must be familiar with how to use a smart phone to be able to utilize this product; persons who are visually impaired or physically challenged (finger/hands) or have problem with other sensory organs of the body will have problem the product [19]. Some of the affected patients may also not have the financial power to afford the cost of maintaining a smart phone and purchasing the smart shoe; thus government intervention will be needed to subsidize the costs. Also, the Bluetooth module used in this work was designed for a maximum range of 10m; thus the smart phone and smart shoe will not work if separated for more than 10m. Bluetooth radio propagation is also susceptible to signal attenuation that can result from obstructions and signal fading that can result from reflections. Another important limitation is that the sensor behavior can vary with environment/climate condition; thus, the sensors will need to be calibrated differently for varying climate conditions.

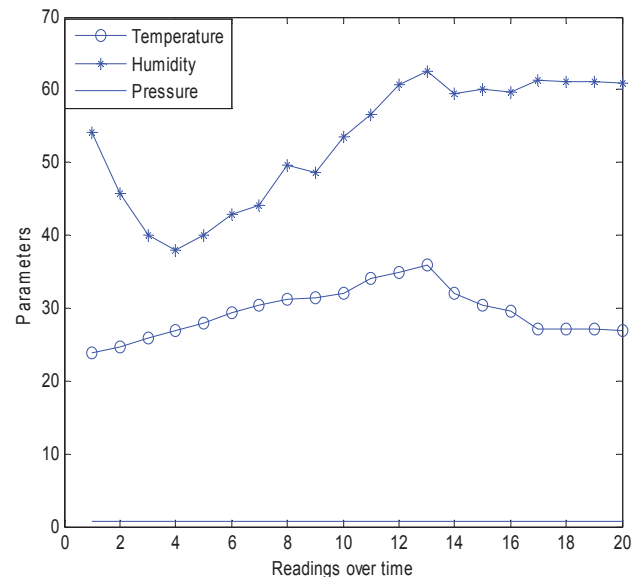


Fig. 6 Parameters' readings with time.

## 6. Conclusion and Future Work

This work is an improvement on our previous work – a computer engineering approach to detect, prevent and manage diabetic foot diseases [10]; it was discussed as future research directions then. We have replaced the non-i2c capable sensors with i2c capable ones (hih6130).

The temperature, humidity and pressure sensors deployed also have smaller sizes, better sensitivity and wider measurement range than the ones that were used previously. The size being small allowed us to fabricate a PCB per hotspots and combine the two sensors on the PCB in a smart and efficient manner. The enhanced capabilities of the sensors result in improved performance of the smart shoe and the system model as a whole. The smart shoe is now able to send sensor readings to the smart phone app in less than a second. Future work is to implement the central server component of the system model and get data transmitted from the phone to the server via LAN or WAN. The central server will run software(s) that can intelligently interpret the data for the purpose of monitoring, analysis and diagnosis.

## 7. Acknowledgements

The authors will like to acknowledge the technical support of Ramin Soltanzadeh, as well the support of Dr. Bob McLeod and Dr. Marcia Friesen of the Internet Innovation center, Department of Electrical and Computer Engineering, University of Manitoba for advice during this research.

## References

- [1] International Diabetic Federation, IDF Diabetes Atlas, 6 ed., 2013.
- [2] Canadian Diabetes Association, "At the tipping point: Diabetes in Manitoba," [Online]. Available: <https://www.diabetes.ca/CDA/media/documents/publications-and-newsletters/advocacy-reports/canada-at-the-tipping-point-manitoba-english.pdf>. [Accessed 21 04 2014].
- [3] D. G. Armstrong, "At-Home Skin Temperature Monitoring Reduces Diabetic Foot Ulceration," *Review of Endocrinology*, pp. 53-55, 2008.
- [4] Orpyx, "SurroSense Rx System," Orpyx, [Online]. Available: <http://orpyx.com/pages/surrosense-rx>. [Accessed 23 04 2014].
- [5] M. Zequera, L. Garavito, W. Sandham, J. C. Bernal, Á. Rodríguez, L. C. Jiménez, A. Hernández, C. Wilches and A. C. Villa, "Diabetic Foot Prevention: Repeatability of the Loran Platform Plantar Pressure and Load Distribution Measurements in Nondiabetic Subjects during Bipedal Standing—A Pilot Study," *Electrical and Computer Engineering*, vol. 2011, pp. 1-14, 2011.
- [6] R. Kanade, R. V. Deursen, K. Harding and P. Price, "Investigation of standing balance in patients with diabetic neuropathy at different stages of foot complications," *Clinical Biomechanics*, vol. 23, no. 9, p. 1183–1191, 2008.
- [7] R. G. Frykberg, A. Tallis and E. Tierney, "Diabetic Foot Self Examination with the Tempstat™ as an Integral Component of a Comprehensive Prevention Program," *The Journal of Diabetic Foot Complications*, vol. 1, no. 1, pp. 13 - 18, 2009.
- [8] L. A. Lavery, K. R. Higgins, D. R. Lanctot, G. P. Constantinides, R. G. Zamorano, D. G. Armstrong, K. A. Athanasiou and C. M. Agrawal, "Home Monitoring of Foot Skin Temperatures to Prevent Ulceration," *Diabetes Care*, vol. 27, no. 11, pp. 2642-2647, 2004.
- [9] Centers for Disease Control and Prevention, "Prepare for diabetes care in heat and emergencies," 1 07 2013. [Online]. Available: <http://www.cdc.gov/features/DiabetesHeatTravel/>. [Accessed 22 04 2014].
- [10] O. D. Jegede, K. Ferens, B. Griffith, B. Podaima and R. Soltanzadeh, "A Computer Engineering Approach to Detect, Prevent and Manage Diabetic Foot Diseases," in *International Conference on Embedded Systems and Applications (ESA) - WorldComp*, Las Vegas, 2014.
- [11] Adafruit, "DHT22 temperature and humidity sensor," [Online]. Available: <http://www.adafruit.com/product/385>. [Accessed 16 02 2015].
- [12] Honeywell Sensing and Control, "http://uk.farnell.com/honeywell-s-c/hih6130-021-001/sensor-humidicon-5-soic-8/dp/1961727," 2014. [Online]. Available: <http://www.farnell.com/datasheets/1765766.pdf>. [Accessed 28 09 2014].
- [13] S. Casagrande, "Firmware for Temp&Humidity Sensor," Galvant, 2014. [Online]. Available: [https://github.com/Galvant/temp\\_sensor\\_firmware](https://github.com/Galvant/temp_sensor_firmware).
- [14] Tekscan, "Thin and Flexible Strain Gauge Alternative," Tekscan, [Online]. Available: <http://www.tekscan.com/strain-gauge>. [Accessed 16 02 2015].
- [15] Tekscan, "FlexiForce® Sensors - Standard FlexiForce Sensors for Force Measurement," 2014. [Online]. Available: <http://www.tekscan.com/flexible-force-sensors>. [Accessed 2014].
- [16] Tekscan, "FlexiForce® Sensors User Manual (Rev H)," Tekscan, South Boston, 2010.

- [17] Sparkfun, "Flexiforce Pressure Sensor (25lbs) Quick Start Guide," Sparkfun, 3 Jan 2013. [Online]. Available: <https://www.sparkfun.com/tutorials/389>.
- [18] Kerimil, "How to control arduino board using an android phone and a bluetooth module," 18 02 2013. [Online]. Available: <http://www.instructables.com/id/How-control-arduino-board-using-an-android-phone-a/>. [Accessed 24 04 2014].
- [19] J. Woodbridge, A. Nahapetian, H. Noshadi, M. Sarrafzadeh and W. Kaiser, "Wireless Health and the Smart Phone Conundrum," in *The 2nd Joint Workshop On High Confidence Medical Devices, Software, and Systems (HCMDSS) and Medical Device Plug-and-Play (MD PnP) Interoperability*, San Francisco, CA, 2009.
- [20] S. Wild, G. Roglic, A. Green, R. Sicree and H. King, "Global Prevalence of Diabetes - Estimates for the year 2000 and projections for 2030," *Diabetes Care*, vol. 27, no. 5, pp. 1047-1053, May 2004.
- [21] G. Danaei, M. M. Finucane, Y. Lu, G. M. Singh, M. J. Cowan, C. J. Paciorek, J. K. Lin, F. Farzadfar, Y.-H. Khang, G. A. Stevens, M. Rao, M. K. Ali, L. M. Riley, C. A. Robinson and M. Ezzati, "National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants," *Lancet*, pp. 1-10, 2011.
- [22] L. E. Dyck, "Diabetes and Aboriginal Canadians," Saskatchewan, 2010.
- [23] Wikipedia, "Peripheral Neuropathy," 2014. [Online]. Available: [http://en.wikipedia.org/wiki/Peripheral\\_neuropathy](http://en.wikipedia.org/wiki/Peripheral_neuropathy). [Accessed 21 04 2014].
- [24] D. Armstrong, M. Sangalang, D. Jolley, F. Maben, H. Kimbriel, B. Nixon and I. Cohen, "Cooling the foot to prevent diabetic foot wounds," *Journal of the American Podiatric Medical Association*, vol. 103, pp. 103-108, 2005.
- [25] Arduino, "Class for DHT11, DHT21 and DHT22," Arduino, 11 02 2014. [Online]. Available: <http://playground.arduino.cc/Main/DHTLib>. [Accessed 24 04 2014].
- [26] C. Arthur, "Nokia revenues slide 24% but Lumia sales rise offers hope," 2013.
- [27] Massachusetts Institute of Technology, "MIT App Inventor," 2014. [Online]. Available: <http://appinventor.mit.edu/explore/>.
- [28] Itead Studio, "Electronic Brick of HC-06 Serial Port Bluetooth. Available:," 2013. [Online]. Available: [http://inmotion.pt/documentation/others/INM-0750/DS\\_IM120710006.pdf](http://inmotion.pt/documentation/others/INM-0750/DS_IM120710006.pdf). [Accessed 24 04 2014].
- [29] Digi-Key Corporation, "CEA-06-250UW-350," Micro-Measurements (Division of Vishay Precision Group), 2014. [Online]. Available: <http://www.digikey.com/product-search/en?vendor=0&keywords=1033-1013-nd>. [Accessed 24 04 2014].
- [30] Government of Manitoba, "Manitoba Healthy Living," [Online]. Available: <http://www.manitobahealthyliving.ca/are-you-at-risk>. [Accessed 21 04 2014].
- [31] P. Fenner, "Reading Strain Gauge Scales with Arduino," Deferred Procrastination, 2014. [Online]. Available: <https://www.deferredprocrastination.co.uk/blog/2013/reading-strain-gauge-scales-with-arduino/>. [Accessed 04 24 2014].
- [32] E. E. S. Exchange, "How to wire up a 3-wire load cell/strain gauge and an amplifier?," Electrical Engineering Stack Exchange, 2012. [Online]. Available: <http://electronics.stackexchange.com/questions/18669/how-to-wire-up-a-3-wire-load-cell-strain-gauge-and-an-amplifier>. [Accessed 24 04 2014].
- [33] Wikipedia, "Bluetooth," Wikipedia, 21 May 2014. [Online]. Available: <http://en.wikipedia.org/wiki/Bluetooth>. [Accessed 23 May 2014].

# Utilizing 3D Printing to Assist the Blind

Rabia Jafri<sup>1</sup> and Syed Abid Ali<sup>2</sup>

<sup>1</sup>Department of Information Technology, King Saud University, Riyadh, Saudi Arabia

<sup>2</sup>Araware LLC, Wilmington, Delaware, U.S.A

**Abstract:** *The rapid advent of 3D printing technologies in recent years has far-reaching implications for individuals with disabilities in general and those with blindness in particular. This paper explores how 3D printing is being utilized for the blind as an affordable, easily accessible means of creating tactile representations of visual information and also expounds upon the possibilities which this technology opens up for blind people, their caregivers and accessibility designers and researchers in terms of enabling them to rapidly and cost-effectively design, create and share assistive tools and components.*

**Keywords:** 3D printing, assistive technologies, blind, visually impaired, tactile representation, accessibility.

## 1. Introduction

The rapid advent in 3D printing technologies over the past decade has resulted in tools capable of producing high quality three-dimensional objects with a variety of customization options in terms of textures, colors and materials (e.g., plastic, wood, metal). These printers are not only becoming available at increasingly affordable prices - due to advancements in technology, expiring patents and increasing competition [1] - but are also being provided with user-friendly interfaces enabling non-experts to easily create custom products to meet their needs. The proliferation of these technologies has far-reaching implications for people suffering from blindness for whom the high prices of assistive solutions - resulting from the increased financial costs involved in the industrial production of limited batches of products tailored to a small user group [2]- coupled with the lack of access to any affordable means of manufacturing customized implements themselves have traditionally been a major hurdle to the procurement of products to aid them in accessing information and performing daily tasks.

This paper aims to provide an overview of how 3D printing is currently being utilized to assist blind individuals from two perspectives: first, by providing them with tactile representations of visual information useful in domains such as education and navigation, and secondly, by enabling them to create and adapt their own assistive devices.

The rest of the paper is organized as follows: Section 2 describes how 3D printing is being utilized to convey visual information to the blind. Section 3 delineates the potential of this technology for allowing blind users, their caregivers and accessibility designers and researchers to inexpensively

create customized assistive solutions. Section 4 points out some online resources currently available to support 3D printing for the blind. Section 5 concludes the paper and indicates some directions for future work.

## 2. 3D printing for conveying visual information

In the absence of the sense of sight, blind individuals mainly rely on their sense of touch and hearing to access information which is available to sighted people in visual form (such as the form, structure and texture of various entities in their environment). Unfortunately, in many cases, much of this information may not be amenable to the sense of touch either because it is presented in two-dimensional form (for example, as images or distant scenery) or because the entities are on a large scale (e.g., buildings) or physically inaccessible precluding tactual exploration. Though a sighted observer may provide verbal descriptions of these details but these are usually not sufficient for blind individuals to construct complete and accurate mental representations of physical entities [3].

Various assistive solutions have been developed to address these problems. Some of these systems decipher images, real-world objects and/or scenes and convey audio descriptions of these either in the form of speech [4-9] (e.g., screen readers [10, 11]) or via sonification (e.g., the vOIce system [12]). However, this kind of output may be quite slow, tedious and incomplete; also, a steep learning curve is required to learn to decipher the sonification output which may be impractical for many blind users [4, 13]. Other solutions relay the descriptions in tactile form either as Braille or by producing tactile images and models. However, Braille descriptions suffer from the limitations outlined for verbal speech descriptions above. Also, until recently, the hardware for producing tactile images and models was prohibitively expensive and not easily accessible to the target users [14, 15].

3D printing offers a cost-effective, accessible means of presenting visual information in tactile form. Two-dimensional images can be converted into computer-aided design (CAD) models of raised outline images or tactile graphics utilizing available and emerging software while three-dimensional objects can be directly scanned using a 3D scanner to generate their 3D CAD models. The CAD models can then be conveniently and inexpensively printed out in physical form using a 3D printer.

3D printing as a means of conveying visual information to the blind is being utilized in several application domains. A few of the most prominent areas in which this technology is being employed are highlighted below:

## 2.1 Navigation

3D printing is being used more and more as a convenient method to generate tactile maps for blind people, with multiple height layers representing various topographic objects (such as roads, paths, water bodies, etc.), to allow them to navigate in both indoor and outdoor spaces. For instance, Gotzelmann and Pavkovic [16] introduce an approach that automatically generates 2.1D tactile models – blind users can customize the level of detail that appears in the model - from the OpenStreetMap [17] data which can then be 3D printed. Schwarzbach [18] created a 3D model for visually impaired users representing the Slovalla recreation area in South Finland using data extracted from aerial photographs and a LIDAR (Light Detection and Ranging) system. Gual et al. [19, 20] conducted a study in Barcelona, Spain with blind and visually impaired individuals to analyze the use and efficacy of tactile maps produced by 3D printing and investigated how the range of tactile symbols used on such maps can be extended. It should be noted that the idea of tactile maps for the visually impaired is not new but the high cost of producing such maps and the inflexibility in terms of customizing them put them beyond the reach of most visually impaired users.

Voigt and Martens [21] propose that 3D printed tactile scale models of the interior architecture of buildings can be used to convey spatial information to blind individuals, thus, serving as an orientation aid for indoor navigation. They suggest placing such models in public buildings in locations that are spatially complex, such as in the vicinity of staircases.

## 2.2 Deciphering the content of digital and printed images

Digital and printed images are used to depict a wide variety of information, e.g., the appearance of objects, people and places, sceneries, graphs, diagrams, etc. Unfortunately, blind people have limited – if any – access to this information and that too, is usually in the form of brief verbal descriptions. However, several techniques are being proposed to convert these 2D images into 2.1D and 3D CAD models which can then be 3D printed to obtain tactile representations of these images. For example, Bajcsy et al. [22] introduce a method for translating web graph and map images into 3D printed models to give the blind and visually impaired access to voting and election data. The Tactile Books Project [23, 24] launched by researchers at The University of Colorado aims to convert basic children's picture books into 3D printed tactile books with 3D figures, sometimes along with Braille text, embedded in the pages (Figure 1). As discussed in the section on navigation, research is also underway on



Figure 1. An example of a 3D printed page from a tactile storybook (image courtesy of [25])

generating 3D printed tactile maps from digital images [16, 18].

Another area being actively explored is the development of methods for producing 3D printed tactile representations of paintings to facilitate the access of visually impaired people to artwork in museums. For example, Volpe et al. [26] propose four alternative computer-based methods for semi-automatic generation of tactile 3D models starting from RGB digital images of paintings. The Midas Touch project [27] at Harvard Innovation Lab also aims to convert paintings into tactile form by adding layers of texture to the 2D images of paintings.

## 2.3 Tactile exploration of remote, huge, miniscule and/or tactually inaccessible objects

Some objects cannot be explored via touch because of their size (they may be too large (e.g., buildings, sculptures, etc.), or too small (e.g., atoms molecules)), their distance (e.g., planets, remotely located objects) and/or their inaccessibility despite their physical proximity (e.g., internal body organs, museum artifacts, zoo animals, etc.). However, since these objects can be visually perceived - in some cases, by using tools like telescopes, microscopes and X-rays, their dimensions can be scanned and measured enabling their 3D CAD models to be produced and 3D printed.

A few examples of how 3D printing has been utilized in this manner are as follows:

**2.3.1 Huge or miniscule objects:** Celani et al. [14] have produced 3D printed tactile scale models of four buildings designed by a renowned architect to convey their architectural details to high school students. The Guide4Blind project [28] has utilized 3D printing to produce scale models of landmark buildings in the city of Soest, Germany. These models were then used to generate bronze models which are placed alongside the landmarks allowing visually impaired travelers to feel them and receive detailed impressions of the shapes and textures of the buildings. A library in Lithuania has also created 3D printed models of several historical buildings (such as the Taj Mahal and Reims Cathedral) and local landmarks for the same purpose out [29] (Figure 2). On the other hand, Teshima et al. [30] have created enlarged 3D



printed skeleton models of micro-organisms for teaching blind pupils about their structure while Wedler et al. [31] have developed a system which allows visually impaired chemistry students to input computed molecular structures which are then printed out in physical form using a 3D printer enabling the students to tactually explore them.



Figure 2. A 3D printed model of the Reims Cathedral [29]

**2.3.2 Remote objects:** The 3D Astronomy Project at the Space Telescope Science Institute and NASA Goddard have created innovative education materials and 3D models of astronomical objects using Hubble data [32]. Since blind people cannot view photographs of celebrities and thus, cannot form an idea of what their faces look like, a library in Lithuania gathered several famous singers and actors, captured their face images from several angles, stitched them together to create 3D models and 3D printed them out [29]. The collection of 3D printed models has been exhibited in several cities.

**2.3.4 Tactually inaccessible objects:** Since blind parents are unable to view ultrasound images of their unborn baby, the Feto3D project [33] employs 3D prenatal imaging technology and 3D printing to produce physical replicas of a baby growing in the womb, allowing moms- and dads-to-be to hold a model of their unborn child in their hands. The same system can also produce 3D printed models of internal body organs and the skeletal system. Yahoo Japan's Hands On Search machine allows visually impaired children to verbally specify objects (e.g., "giraffe" or "Tokyo Skytree Building") [34]. It then searches Yahoo and upon finding the object, 3D prints out a miniaturized version of it for the child to tactually explore. Neumüller et al. [15] have expounded upon the potential of 3D printing as a means for making cultural artifacts in museums accessible to visually impaired people and have provided several examples of this [35, 36].

## 2.4 Education

Physical models are considered to be indispensable tools for teaching visually impaired children, especially those with congenital blindness, to help them form concepts of things

that they have never seen [14, 15, 37]. However, the high expense of such models as well as the decreasing costs of simulation and visualization software and computer modelling have increasingly deterred educators from utilizing such models in their teaching [15]. 3D printing as a low cost option for producing physical models can turn that situation around. 3D printing for deciphering the contents of digital and printed images and the tactile exploration of remote, huge, miniscule and/or tactually inaccessible objects, as discussed in the preceding sections, has obvious applications in education, with several projects aimed specifically towards children's education, e.g., the Tactile Books Project [23, 24], Yahoo Japan's Hands On Search machine, NASA's 3D Astronomy Project [32], and the systems for teaching about molecular structures [31], microorganisms [30] and architecture [14] mentioned above. Some additional examples include the use of 3D printed models to convey chemistry and mathematics concepts to visually impaired students at Simon Fraser University, Canada [38], converging 3D printing and 3D surface thermal reflow treatment techniques to produce touchable objects with detailed lines and curves at the Korea Institute of Science and Technology [39] and having visually impaired pupils write code to produce accessible 3D printed tactile visualizations of data extracted from Twitter at the National Federation for the Blind's STEM-X camp [40].

3D printed models can also be incorporated into educational software for children. For example, we are currently implementing a system for teaching tactual shape perception and small scale space spatial relationships to visually impaired children which utilizes a tangible user interface consisting of 3D printed objects in various shapes and geometrics forms which can be identified and tracked by a computer vision based system as they are moved across a transparent surface [41].

It should be noted that 3D printed models can not only convey visual information to a student, they can also enable the student to relay information from his perspective to others in a form that is accessible to him as described in [31].

## 3. 3D printing for creating customized assistive devices

People with disabilities frequently rely on a variety of assistive devices to help them carry out daily activities. Though many such devices are commercially available, several studies have shown that about one third of all such devices are eventually abandoned by their users [42-44]. Some reasons identified for this high rejection rate include the following [45]: Assistive technologies (AT), which have generic designs, cannot meet the complex and ever-changing needs of people with disabilities who often suffer from multiple conditions [2, 46, 47]. The lack of user involvement in the selection of AT [42, 47, 48] and the often-observed proclivity of the disabled to purchase assistive devices just because they are easy to procure even though they do not meet their individual needs [42] have been cited as other

significant factors leading to abandonment. Moreover, the high prices of AT, arising from the high development costs of producing small batches of products tailored for small user groups [2], put them beyond the financial reach of most disabled people and their caregivers.

3D printing offers a viable solution to all the above problems [2, 49, 50]. It allows blind users and their caregivers to design and print out their own assistive solutions customized to their unique needs and at a fraction of the price that it would cost to have the same solutions custom manufactured by a retailer. Moreover, if the user's needs change, the designs can be adapted and/or new tools can be created to meet the user's new requirements. With 3D printers becoming increasingly accessible and affordable, users can receive their assistive solutions at much faster speeds and with greater ease as compared to the often complex and lengthy procedures required to procure the same devices via referrals to multiple, often uncoordinated, organizations [49, 51, 52]. Furthermore, if a part gets broken or lost, it can be conveniently replaced.

Moreover, 3D printing allows researchers working in the area of assistive technologies for the blind to custom manufacture several of the hardware components for their prototypes. This offers two main advantages: 1. The CAD models of these components can be provided to the users enabling them to print these out and put together the prototype devices themselves. 2. It allows for the creation of a more aesthetically pleasing solution than one produced by clumsily attaching sensors and motors to existing products resulting in bulky, unwieldy, outlandish devices (this is especially significant since, as reported by several studies [53], the "cosmetic acceptability" of a device is perceived to be more important by visually impaired users than the actual functions it provides.).

A few examples of how 3D printing has been utilized for building assistive devices for the blind are as follows:

Several educational toys and games are being developed for VI children whose components can be 3D printed. For instance, Fittle [54] is a set of 3D printed puzzles where each puzzle consists of blocks imprinted with Braille letters which, when put together correctly, not only form a word but also the shape of the object that the word represents (e.g., blocks with letters "F", "I", "S" and "H" can be put together to form the shape of a fish (Figure 3)). This allows a child to learn both how to spell the word and also to visualize it. 3D printed objects can also be used to provide a tangible user interface to educational software. For instance, we proposed a solution which teaches Braille letter recognition to young blind children by allowing them to manipulate NFC-tag embedded blocks with Braille letters embossed on them [55]. The blocks can be 3D printed and customized according to individual needs. Children can interact with the system by providing input via the tangible interface and receive auditory feedback via a speech-based interface. Another system for teaching tactual shape perception and spatial relationships to blind children via a tangible user interface consisting of 3D printed

geometric objects has already been described in section 2.4 [41].



**Figure 3. Example of a Fittle puzzle [54].**

Kane et al. [56] present touchplates, tactile guides which may be 3D printed, among other options, and overlaid on touchscreens as an aid for providing tactile feedback to blind users.

Nanayakkara et al. [57] introduce EyeRing, a 3D printed ring with an embedded camera, worn on the finger. The user can point the EyeRing towards an area of interest and press a button on the side of the ring to capture an image which is then transmitted via Bluetooth to a mobile phone application which analyzes the image and provides feedback to the user via a text-to-speech module. Some prototypical applications using this device for navigation, currency detection and color detection were tested. EyeRing has also been expanded to the FingerReader device [58] for reading printed text (Figure 4).



**Figure 4. The FingerReader device [58] for reading printed text (image courtesy of [59])**

Similarly, researchers at Oxford University [60] are developing glasses whose frames can be 3D printed and then equipped with various sensors and electronic components to enable people with severely impaired sight to use their remaining vision to detect obstacles (Figure 5). Furthermore, several commercial assistive products for the blind are appearing in the market, such as the Braille phones developed by OwnFone [61], which are incorporating 3D printed components to keep their cost down.



**Figure 5. Experimental prototype of 'smart glasses' to enhance vision for poorly sighted individuals. This system has a see-through display and a special type of camera that can detect and highlight nearby objects [60].**

#### 4. 3D printing resources for the blind

It should be noted that a number of resources are available online to support 3D printing for the blind. For instance, the Accessibility Metadata Project [62], aims to add metadata tags to the digital files for 3D printing available on MakerBot's Thingsiverse website [63], to enable search engines to pick out only objects which are tactile. Individual users (e.g., [64], [65]) have also uploaded their designs for 3Dprintable products for blind individuals in various online communities.

Various user friendly tools for converting 2D data into 3D models are also being developed. Some have already been mentioned in section 2.2. Another example is VizTouch [66], a software, developed through a user-centered approach involving visually impaired users. It allows a user to input an equation or an excel file for the information which he would like to visualize and automatically generates an STL file ready to be printed on a 3D printer. VizTouch has been equipped with a screen reader to allow visually impaired users to access it.

Tools like the Easy Make Oven [49] allow users with little or no 3D modelling experience to scan, alter and combine physical objects and then export them in a format ready for fabrication on a 3D printer or laser cutter.

In addition, there are several libraries of 3D models available online which, while not specifically targeted towards blind users, can be 3D printed to offer tactile representations of various kinds of physical objects and visual information [63] [37].

#### 4. Conclusion

This paper explored how 3D printing is being utilized for the blind as an affordable, easily accessible means of creating tactile representations of visual information. It also expounded upon the possibilities which this technology opens up for blind people, their caregivers and accessibility designers and researchers in terms of enabling them to rapidly and cost-effectively design, create and share assistive

tools and components. Moreover, it listed some online resources currently available for this purpose.

The continuing proliferation of 3D printers with decreasing prices and increasing functionalities and options, and the ensuing development of more user-friendly 3D modelling software tools coupled with the online global sharing of ideas and resources indicate that the potential of this technology will continue to expand. We are eager to see how this potential will be harnessed in innovative ways to assist blind people in their everyday lives.

#### References

- [1] A. Shah. (2014, April 10, 2015). 3D printer price drops could lure home users. Available: <http://www.pcworld.com/article/2140360/3d-printer-price-drops-could-lure-home-users.html>
- [2] L. De Couvreur and R. Goossens, "Design for (every)one : co-creation as a bridge between universal design and rehabilitation engineering," *CoDesign*, vol. 7, pp. 107-121, 2011.
- [3] Saskatchewan. Department of Learning. Special Education Unit, *Teaching Students with Visual Impairments: A guide for the support team*: Saskatchewan Learning, 2003.
- [4] R. Jafri, S. Ali, H. Arabnia, and S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: a survey," *The Visual Computer*, vol. 30, pp. 1197-1222, 2014/11/01 2014.
- [5] R. Jafri, S. A. Ali, and H. R. Arabnia, "Computer Vision-based object recognition for the visually impaired using visual tags," in *Proceedings of the 2013 International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV '13)*, Las Vegas, Nevada, USA, 2013, pp. 400-406.
- [6] R. Jafri and S. A. Ali, "A Multimodal Tablet-based Application for the Visually Impaired for Detecting and Recognizing Objects in a Home Environment," in *Proceedings of the 14th International Conference on Computers Helping People with Special Needs (ICCHP 2014)*, Paris, France, 2014, Springer International Publishing, *Lecture Notes in Computer Science (LNCS)*, Volume 8547, pp. 356-359.
- [7] R. Jafri, S. A. Ali, and H. R. Arabnia, "Face recognition for the visually impaired," in *Proceedings of the 2013 International Conference on Information and Knowledge Engineering (IKE '13)*, Las Vegas, Nevada, USA, 2013, pp. 153-159.
- [8] R. Jafri and S. A. Ali, "Exploring the Potential of Eyewear-Based Wearable Display Devices for Use by the Visually Impaired," in *Proceedings of the 3rd International Conference on User Science and Engineering (i-USER 2014)*, Shah Alam, Malaysia, 2014, pp. 119-124.
- [9] R. Jafri and S. A. Ali, "A GPS-based Personalized Pedestrian Route Recording Smartphone Application for the Blind," in *Proceedings of the 16th International Conference on Human-Computer Interaction (HCI 2014) - Posters' Extended Abstracts*, Crete, Greece, 2014, Springer

- International Publishing, *Communications in Computer and Information Science (CCIS)*, Volume 435, pp. 232-237.
- [10] Y. Borodin, J. P. Bigham, G. Dausch, and I. V. Ramakrishnan, "More than meets the eye: a survey of screen-reader browsing strategies," in *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, Raleigh, North Carolina, 2010, pp. 1-10.
- [11] "JAWS Screen reader, Freedom Scientific, Inc. <http://www.freedomscientific.com/Products/Blindness/JAWSh>."
- [12] P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, pp. 112-121, February 1992.
- [13] W. Ugulino and H. Fuks, "Prototyping Wearables for Supporting Cognitive Mapping by the Blind: Lessons from Co-Creation Workshops," in *Proc. of the 2015 workshop on Wearable Systems and Applications*, 2015, pp. 39-44.
- [14] G. Celani, V. Zattera, M. de Oliveira, and J. da Silva, "'Seeing' with the Hands: Teaching Architecture for the Visually-Impaired with Digitally-Fabricated Scale Models," in *Global Design and Local Materialization*. vol. 369, J. Zhang and C. Sun, Eds., ed: Springer Berlin Heidelberg, 2013, pp. 159-166.
- [15] M. Neumüller, A. Reichinger, F. Rist, and C. Kern, "3D Printing for Cultural Heritage: Preservation, Accessibility, Research and Education," in *3D Research Challenges in Cultural Heritage*. vol. 8355, M. Ioannides and E. Quak, Eds., ed: Springer Berlin Heidelberg, 2014, pp. 119-134.
- [16] T. Götzelmann and A. Pavkovic, "Towards Automatically Generated Tactile Detail Maps by 3D Printers for Blind Persons," in *Computers Helping People with Special Needs*. vol. 8548, K. Miesenberger, D. Fels, D. Archambault, P. Peñáz, and W. Zagler, Eds., ed: Springer International Publishing, 2014, pp. 1-7.
- [17] M. Haklay and P. Weber, "OpenStreetMap: User-Generated Street Maps," *Pervasive Computing, IEEE*, vol. 7, pp. 12-18, 2008.
- [18] F. Schwarzbach, T. Sarjakoski, J. Oksanen, L. T. Sarjakoski, and S. Weckman, "Physical 3D models from LIDAR data as tactile maps for visually impaired persons," in *True-3D in Cartography*, M. Buchroithner, Ed., ed: Springer Berlin Heidelberg, 2012, pp. 169-183.
- [19] J. Gual, M. Puyuelo, J. Lloverás, and L. Merino, "Visual Impairment and urban orientation. Pilot study with tactile maps produced through 3D Printing," *Psychology*, vol. 3, pp. 239-250, 2012.
- [20] J. Gual, M. Puyuelo, and J. Lloveras, "Analysis of volumetric tactile symbols produced with 3D printing," in *ACHI 2012, The Fifth International Conference on Advances in Computer-Human Interactions*, 2012, pp. 60-67.
- [21] A. Voigt and B. Martens, *Development of 3D tactile models for the partially sighted to facilitate spatial orientation*. na, 2006.
- [22] A. Bajcsy, Y.-S. Li-Baboud, and M. Brady, "Electronic Imaging & Signal Processing Depicting Web images for the blind and visually impaired."
- [23] A. Stangl, J. Kim, and T. Yeh, "3D printed tactile picture books for children with visual impairments: a design probe," in *Proceedings of the 2014 conference on Interaction design and children*, Aarhus, Denmark, 2014, pp. 321-324.
- [24] J. Kim, H. Oh, and T. Yeh, "A Study to Empower Children to Design Movable Tactile Pictures for Children with Visual Impairments," in *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, Stanford, California, USA, 2015, pp. 703-708.
- [25] "<https://vimeo.com/106853836>," ed.
- [26] Y. Volpe, R. Furferi, L. Governi, and G. Tennirelli, "Computer-based methodologies for semi-automatic 3D model generation from paintings," *International Journal of Computer Aided Engineering and Technology*, vol. 6, pp. 88-112, 2014.
- [27] J. Vesanto. (2013, April 17, 2015). Midas Touch – Augmented Art Project for the Visually Impaired. *3D Printing Industry*. Available: <http://3dprintingindustry.com/2013/04/19/midas-touch-augmented-art-project-for-the-visually-impaired/>
- [28] (2013, April 30, 2015). Sightseeing by touch: experience the world with 3D models. Available: <http://www.3ders.org/articles/20130521-sightseeing-by-touch-experience-the-world-with-3d-models.html>
- [29] S. J. Grunewald. (2014, April 23, 2015). Steve Wozniak and Lithuanian Celebrities Help the Blind See with 3D Printed Models. *3D Printing Industry*. Available: <http://3dprintingindustry.com/2014/08/01/steve-wozniak-lithuanian-celebrities-help-blind-see-3d-printed-models/>
- [30] Y. Teshima, A. Matsuoka, M. Fujiyoshi, Y. Ikegami, T. Kaneko, S. Oouchi, Y. Watanabe, and K. Yamazawa, "Enlarged Skeleton Models of Plankton for Tactile Teaching," in *Computers Helping People with Special Needs*. vol. 6180, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds., ed: Springer Berlin Heidelberg, 2010, pp. 523-526.
- [31] H. B. Wedler, S. R. Cohen, R. L. Davis, J. G. Harrison, M. R. Siebert, D. Willenbring, C. S. Hamann, J. T. Shaw, and D. J. Tantillo, "Applied Computational Chemistry for the Blind and Visually Impaired," *Journal of Chemical Education*, vol. 89, pp. 1400-1404, 2012/10/09 2012.
- [32] "3D Astronomy: Modelling the Universe with 3D Printers. HubbleSite, NASA. Available: [http://hubblesite.org/get\\_involved/hubble\\_hangouts](http://hubblesite.org/get_involved/hubble_hangouts)," ed.
- [33] T. Baklinski. (2014, April 23, 2015). 3D printing lets blind parents hold model of their unborn child. *Lifesite News*. Available: <https://www.lifesitenews.com/news/3d-printing-lets-blind-parents-hold-model-of-their-unborn-child>
- [34] C. Daileda. (2013, April 23, 2015). Yahoo Japan's 3D Printer Helps Blind Children Search the Web. *Mashable*. Available: <http://mashable.com/2013/09/30/yahoo-japan-blind-3d-printer/>
- [35] N. Gangjee, H. Lipson, and D. I. Owen, "3D Printing of Cuneiform Tablets Cornell Creative Machines Lab, <http://creativemachines.cornell.edu/cuneiform>".
- [36] M. Garber. (2013, The Uncanny Face Model They Made With Richard III's Skull... And it was made by, yep, a 3D printer. *The Atlantic*. Available: <http://www.theatlantic.com/technology/archive/2013/05/the-uncanny-face-model-they-made-with-richard-iiis-skull/275965/>

- [37]M. E. Knapp, R. Wolff, and H. Lipson, "Developing printable content: A repository for printable teaching models," in *Proceedings of the 19th Annual Solid Freeform Fabrication Symposium, Austin TX, USA, 2008*.
- [38]J. Toal, "Modelling flexible learning." <http://flexed.sfu.ca/?p=94>," ed. Simon Fraser University, Canada, 2014.
- [39]"3D Printers Provide Educational Aid to the Visually Impaired." <http://www.engineering.com/3DPrinting/3DPrintingArticles/ArticleID/8038/3D-Printers-Provide-Educational-Aid-to-the-Visually-Impaired.aspx>," ed. Engineering.com, 2014.
- [40]S. K. Kane and J. P. Bigham, "Tracking @stemxcomet: teaching programming to blind students via 3D printing, crisis management, and twitter," in *Proceedings of the 45th ACM technical symposium on Computer science education*, Atlanta, Georgia, USA, 2014, pp. 247-252.
- [41]R. Jafri, A. M. Aljuhani, and S. A. Ali, "A Tangible Interface-based Application for Teaching Tactual Shape Perception and Spatial Awareness Sub-Concepts to Visually Impaired Children," in *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences*, Las Vegas, Nevada, 2015.
- [42]B. Phillips and H. Zhao, "Predictors of Assistive Technology Abandonment," *Assistive Technology*, vol. 5, pp. 36-45, 1993/06/30 1993.
- [43]M. L. Riemer-Reiss and R. R. Wacker, "Factors associated with assistive technology discontinuance among individuals with disabilities," *The Journal of Rehabilitation*, vol. 66, pp. 44-50, 2000.
- [44]R. Verza, M. L. Carvalho, M. Battaglia, and M. M. Uccelli, "An interdisciplinary approach to evaluating the need for assistive technology reduces equipment abandonment," *Multiple sclerosis*, vol. 12, pp. 88-93, 2006.
- [45]J. Hook, S. Verbaan, A. Durrant, P. Olivier, and P. Wright, "A study of the challenges related to DIY assistive technology in the context of children with disabilities," in *Proceedings of the 2014 conference on Designing interactive systems*, Vancouver, BC, Canada, 2014, pp. 597-606.
- [46]A. Kintsch and R. DePaula, "A framework for the adoption of assistive technology," *SWAAAC 2002: Supporting Learning Through Assistive Technology*, pp. 1-10, 2002.
- [47]M. J. Scherer, "The change in emphasis from people to person: introduction to the special issue on Assistive Technology," *Disability & Rehabilitation*, vol. 24, pp. 1-4, 2002.
- [48]M. Dawe, "Desperately seeking simplicity: how young adults with cognitive disabilities and their families adopt assistive technologies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montreal, Quebec, Canada, 2006, pp. 1143-1152.
- [49]A. Hurst and S. Kane, "Making 'making' accessible," in *Proceedings of the 12th International Conference on Interaction Design and Children*, New York, New York, USA, 2013, pp. 635-638.
- [50]A. Hurst and J. Tobias, "Empowering individuals with do-it-yourself assistive technology," in *Proceedings of the 13th International ACM SIGACCESS conference on Computers and accessibility*, Dundee, Scotland, UK, 2011, pp. 11-18.
- [51]J. Copley and J. Ziviani, "Barriers to the use of assistive technology for children with multiple disabilities," *Occupational Therapy International*, vol. 11, pp. 229-243, 2004.
- [52]D. M. Cowan and A. R. Turner-Smith, "The user's perspective on the provision of electronic assistive technology: Equipped for life?," *The British Journal of Occupational Therapy*, vol. 62, pp. 2-6, 1999.
- [53]R. Golledge, R. Klatzky, J. Loomis, and J. Marston, "Stated preferences for components of a personal guidance system for nonvisual navigation," *Journal of Visual Impairment & Blindness*, vol. 98, pp. 135-147, 2004.
- [54]"Fittle." <http://www.fittle.in/>."
- [55]R. Jafri, "Electronic Braille Blocks: A Tangible Interface-based Application for Teaching Braille Letter Recognition to Very Young Blind Children," in *Proceedings of the 14th International Conference on Computers Helping People with Special Needs (ICCHP 2014)*, Paris, France, 2014, Springer International Publishing, *Lecture Notes in Computer Science (LNCS)*, Volume 8548, pp. 551-558.
- [56]S. K. Kane, M. R. Morris, and J. O. Wobbrock, "Touchplates: low-cost tactile overlays for visually impaired touch screen users," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, Bellevue, Washington, 2013, pp. 1-8.
- [57]S. Nanayakkara, R. Shilkrot, and P. Maes, "EyeRing: an eye on a finger," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, Austin, Texas, USA, 2012, pp. 1047-1050.
- [58]R. Shilkrot, J. Huber, W. M. Ee, P. Maes, and S. C. Nanayakkara, "FingerReader: A Wearable Device to Explore Printed Text on the Go," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, 2015, pp. 2363-2372.
- [59]J. Luimstra. (2014, May 27, 2015). "This Ring Enables Visually Impaired to Read". Available: <http://3dprinting.com/news/ring-enables-visually-impaired-person-read/>
- [60]"Oxford Smart Specs Research Group." <http://www.eye.ox.ac.uk/research/oxford-smart-specs-research-group>," ed.
- [61](2014, April 29, 2015). Braille phone goes on sale in 'world first'. *BBC News*. Available: <http://www.bbc.com/news/technology-27437770>
- [62](April 29). *Accessibility Metadata Project: MakerBot Thingiverse 3D Printed Objects*. Available: <http://www.a11ymetadata.org/draft-content/makerbot-thingiverse-3d-printed-objects/>
- [63](April 29). *MakerBot Thingiverse*. Available: <http://www.thingiverse.com/>
- [64]G. Knape. April 29, 2015). *MakerBot Thingiverse Collections*. Available: <http://www.thingiverse.com/knape/collections/>
- [65]April 29, 2015). *GrabCAD: BlindDesign*. Available: [https://grabcad.com/library?per\\_page=20&query=blindesign](https://grabcad.com/library?per_page=20&query=blindesign)
- [66]C. Brown and A. Hurst, "VizTouch: automatically generated tactile visualizations of coordinate spaces," in *Proc. Sixth Int'l Conf. on Tangible, Embedded and Embodied Interaction*, Kingston, Ontario, Canada, 2012, pp. 131-138.

# Changes to User Learning Behaviour of powered wheelchair drivers depending on the level of sensor support

D.A. Sanders *FIET FIMechE FHEA, Member, IEEE*, N Bausch, *MIET*, D. Ndzi, *MIET*

**Abstract** — This paper describes early results in ongoing work to evaluate the effect of intelligent sensor support while a user learns and develops the skill to drive a powered wheelchair. Dependence on training procedures was measured during situations when different levels of support were provided by sensor systems. Results from experiments are presented and some insights are made concerning user learning behaviour during wheelchair driving.

## I. INTRODUCTION

Research into powered wheelchairs is attempting to improve safety and performance and the distribution of tasks and control between human users and powered wheelchair systems is a key issue to be considered [1],[2]. The appropriate level of control and automation is influenced by a variety of factors: level of ability and expertise, mental workload; effectiveness and reliability of the automation; and the users' trust in the automation [3]. It has been reported that the burdens associated with managing automation can sometimes outweigh the potential benefit of the automation to improve system performance. For example, Kirlik studied the interaction between human users and automated systems to investigate why aids might sometimes go unused [4]. Situation Adaptive Autonomy was proposed in which the importance of a change in the level of automation according to dynamically changing situations has been emphasized [5]. It was reported that the time taken to complete a task with a wheelchair partly depends on how the human user interacts with the powered-wheelchair [6],[7]. Other recent work reported that users tend to rely heavily on visual feedback if it is available [3][8],[9] and that the amount of sensor support should be varied depending on circumstances [10],[11].

In the research presented in this paper, the way that users adapted their behaviour in the face of different levels of support is examined. The results are used to evaluate the effect of providing intelligent support during teaching as the users learn and develop their skills. The appropriate level of automation and assistance depends on the complex interaction of several factors and this work investigated how powered wheelchair users should be trained if different levels of support were

available. Recent work had shown that in some circumstances, a skilled user who was trained without any sensor support could sometimes perform even better by using a sensor system to assist them [6]-[11]. The question addressed in this paper was, could a powered wheelchair user trained to achieve tasks with a sensor system to assist them handle a situation without a sensor system? The adaptation behaviour of the users when provided with different levels of support was investigated and it was found that they behaved differently when they encountered different working conditions.

In conventional studies to examine the use of different levels of support, experts with developed skills were taken as examples. The process of acquiring skill has usually been neglected [1]. The task of driving a powered wheelchair was selected as a practical example of human computer interaction and two sets of experiments are described. Similar experiments using remote industrial robot manipulators can be found in the literature [12],[13].

## II. THE POWERED WHEELCHAIR SYSTEM

The apparatus consisted of a dedicated controller with analogue interfacing, DC servo-amplifiers and joystick, and a BobCat II powered wheelchair was modified to include extra control and sensor systems. Two driven wheels were at the front over each driving wheel and two trailing castors at the back. Ultrasonic sensor pairs were mounted over each driving wheel. Altering the differential of rotational speed of the driving wheels affected steering and direction of movement.

Sonar sensors have been widely used for powered-wheelchairs and mobile robots [14],[15] and ultrasonic ranging was selected, as it was simple, cost effective and robust. Ultrasonic transmitter and receiver pairs were mounted at the front of the powered-wheelchair. With suitable processing the ultrasonic signals were converted to a simple representation of the environment ahead of the wheelchair. An integral function was used with the joystick signals so that the tendency to turn when approaching an object could be over-ruled by the user, for example to reach a light switch on a wall.

Software algorithms to intelligently mix the inputs to the powered wheelchair (joystick and sensors) were described in [16]-[20] and the wheelchair was driven under computer control by "fly-by-wire". The direct link between the powered wheelchair and joystick was severed and a computer processed control information. Sensors were activated and interrogated by the computer and the computer was programmed to modify the powered-wheelchair path. Alternatively, joystick control data

D.A. Sanders is a Leverhulme Trust Senior Research Fellow with the Royal Academy of Engineering based at University of Portsmouth, Anglesea, Portsmouth, PO1 3DJ, UK (e-mail: david.sanders@port.ac.uk).

N. Bausch is a Lecturer at University of Portsmouth, Anglesea, Portsmouth, PO1 3DJ, UK (e-mail: nils.bausch@port.ac.uk).

D. Ndzi is a Principal Lecturer at the University of Portsmouth, Anglesea, Portsmouth, PO1 3DJ, UK (e-mail: david.ndzi@port.ac.uk).

could be processed and sent to the wheelchair controller without modification. In this case the powered-wheelchair responded to joystick inputs as if it was an unmodified wheelchair system. Software systems were constructed using methods discussed in [21]-[23]. Systems had three main levels: supervisory, strategic and servo control. These were similar to the levels and sensor systems described or used in [24]-[26].

Algorithms applied the following rules: (1) The user remained in overall control. (2) Systems only modified the trajectory of the powered-wheelchair when necessary. (3) Movements of the wheelchair were smooth and controlled.

### III. EXPERIMENTS

Three levels of support were given to the powered wheelchair drivers:

- Level 0: The ultrasonic sensor system was switched off. The user could steer the manipulator without any disturbance from the automated systems. The user had the most freedom of action in this case but risk of collision was the highest.
- Level 1: The sensor system was switched on and a repulsive force was provided when the powered wheelchair was driven close to obstacles. The magnitude of the repulsive force was inversely proportional to the distance between the obstacle and a sensor.
- Level 2: The sensor system was switched on and the system automatically steered the powered wheelchair away from obstacles. If the driver tried to move the powered wheelchair towards an obstacle, the system automatically steered the powered wheelchair away from the obstacle.

In each case, a driver could move the powered wheelchair against the applied force if they made more effort. The average time to complete a task (T) and the number of collisions (C) during that task were used as measures of performance.

#### A. The first set of experiments

A first set of experiments made a comparison between learning in Level 0 (with the ultrasonic sensor system switched off) and Level 2 (with the sensor system switched on and the system automatically steering the powered wheelchair away from obstacles). A second set of experiments compared learning in Level 0, Level 1 and Level 2 when the environment changed.

In each of four experiments, volunteers were tasked with driving a Bobcat II Powered wheelchair [7], [10] through one of four different courses. The first route was a simple route from a start line between some double doors, along a corridor avoiding three obstacles and then turning a corner to a finish line.

The four courses were of various lengths and different numbers of obstacles were located within the path of the powered wheelchair. Volunteers were instructed to drive the powered wheelchair through each course and to avoid hitting the obstacles. The powered wheelchair drivers used a joystick connected to the powered wheelchair.

For the tests at Level 2, resistive force became stronger if the driver moved the powered wheelchair closer to obstacles. Thus the support was more restrictive compared with Level 1.

Volunteers were sixty University students (without any previous experience). They were divided into two groups (Group A and Group B). Group A and Group B were then divided between the four different courses used for the experiments; roughly eight volunteers in each sub-group (A1, A2, A3, A4, B1, B2, B3, B4). Tests for each of the courses (one to four) took place on different days.

Volunteers were shown their route for the powered wheelchair and the obstacles along the route. Subjects in Group A and Group B performed the task ten times with and without the sensor systems to assist them respectively. Then, subjects in Group A and Group B performed the task with different support conditions. The purpose of second set of experiments was to examine performance when an user encountered new support conditions after they had already developed some skill in driving the powered wheelchair. A simple questionnaire was used to collect subjective information about preferences. Questions asked were: (1) "Which do you prefer Level 0 support or Level 2 support?" and (2) "Which do you think easier to drive; Level 0 support or Level 2 support?"

Responses are shown in Table 1.

#### B. Results from the first set of experiments

Figs 1 and 2 show the average time T of each of Group A and Group B for each attempt at driving along the four different courses. Trial numbers 1 to 10 correspond to the first half (Group A with Level 2 support and Group B with Level 0 support) and 11 to 20 correspond to the second half of the experiments (Group A with Level 0 support and Group B with Level 2 support).

In the first half, subjects in Group A reached the learning equilibrium earlier and more stably than subjects in Group B. This result suggests a positive effect of using Level 2 support during the earlier stages of learning and skill development. Better performance was also observed for Group A during the second half of the trials, when subjects performed the task with Level 0 support. Skill acquisition was accelerated using Level 2 support and subjects may have acquired the general skills which can be applied to the conditions without any support. On the contrary, the performance of subjects in Group B in the second half of the trials did not show any significant improvement in terms of average time and stability even if the Level 2 support was applied.

TABLE I. RESPONSES FROM QUESTIONNAIRES

	Strongly prefer 0	Prefer Level 0	Un-decided	Prefer Level 2	Strongly prefer 2
Preference	13	15	17	12	3
How easy?	14	17	8	15	6

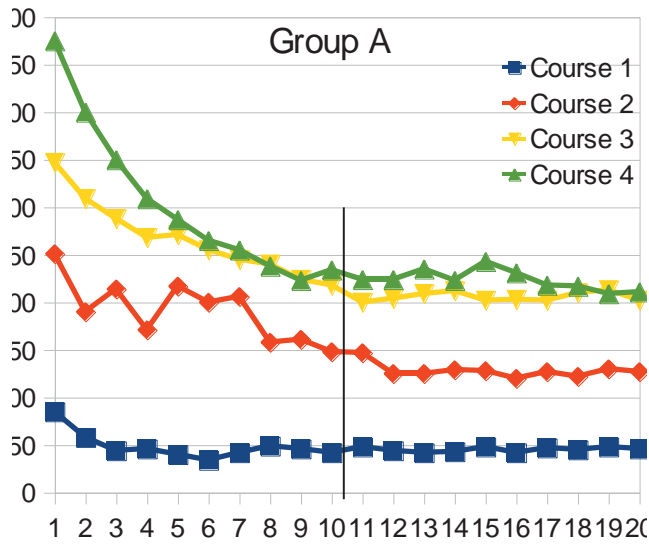


Figure 1. Average time T for Group A to complete four different courses. Level 2 support first, then Level 0

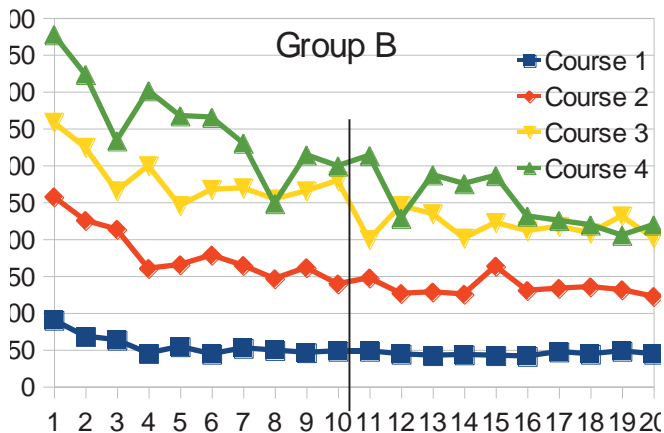


Figure 2. Average time T for Group Group B to complete four different courses. Level 0 support first then Level 2

That behaviour can be attributed to the fact that skill developed during Level 0 support did not transfer so easily to the skills required in a different support condition. The results of the subjective evaluation obtained by questionnaire are shown in Table 1. Subjects suggested that they thought it was easier to drive without support (Level 0) which contradicted the objective results shown above. The subjective evaluations also implied that constraint-based support was disliked.

### C. Discussion of the first set of experiments

Results from the first set of experiments suggested that a driver who had been trained with a sensor system supporting them could still handle situations when the systems were removed. The support from the sensor systems during the training phase also had a positive effect on performance without any support. It should be noted though that subjects may have become accustomed to maneuvering the powered wheelchair

through the same path during the sequence of trials. Further experiments will be required to confirm that the developed skill can be transferred to different working conditions without the support function. The results agree with those suggested by Chikura [1] but the behaviour of subjects facing different types of task needs to be examined if results are to be confirmed and generalized.

Results suggested that a driver trained without the support of the sensor systems could not perform better using the support of the sensor systems. This may be because the task needed motor skill. Results indicate that subjects trained without support did not show steady learning compared with subjects who started without any previous experience. This suggests that skill gained when driving without any assistance had a negative effect on the performance with the support function. This is important when considering a training procedure. Informal interviews with subjects revealed that this was partly because of a feeling that freedom of movement was being constrained. This inconsistency between the performance results and the subjective evaluation indicated that selecting options based on human preference could lead to inferior performance.

### D. Second set of experiments

A second set of experiments is now being conducted to investigate adaptive behaviour when working conditions were changed and volunteers were using different levels of support. A new route is being used in the second set of experiments that is longer and more complicated. That is allowing obstacles to be moved to create three different (but similar) courses along the same route. Progress so far is described.

The purpose of the experiments in phase two was to investigate whether there were any differences in subjects' behaviour for different levels of support when they encountered different working conditions. First, the route was completed six times. Then the route was modified by moving the obstacles for both the second set and then the third set of six attempts to create two new different courses.

Each driver performed each task with the same support conditions throughout the three sets of six attempts over each of the three courses. Ten university students without previous experience (they had not participated in the first set of experiments), were divided into three groups (Groups X, Y and Z). Subjects in Group X performed the task with level 0 support (manual control) and Group Y with level 1 support (sensor system providing a repulsive force) and Group Z with Level 2 support (sensor system automatically steering the powered wheelchair away from obstacles). Task completion time T and number of collisions C were recorded.

### E. Results from the second set of experiments

The resultant T for the second set of experiments for Groups X, Y and Z respectively and the number of collisions with obstacles for each subject are being recorded. So far, it appears that Level 2 support results in the lowest number of collisions. Contrary to this positive effect though, the performance in terms of T degraded compared with Level 0 support. To examine the



net effect of each support Level, the differences in T when the configuration of obstacles changed was evaluated. T tended to increase in cases of Level 1 and 2 support, while T did not show any significant change in case of Level 0 support. This result indicates that skills acquired without any support may be more general compared with skills obtained with support.

#### F. Discussion of the second set of experiments

There do appear to be differences in adaptation and behaviour for different levels of support when drivers encounter a new and different working condition. T is showing different behaviour for each support level when the configuration of the obstacles was changed.

### IV. OVERALL DISCUSSION AND CONCLUSIONS

Adaptation behaviour to different levels of support was examined to evaluate the effect of using systems to assist during teaching and learning and skill development. The positive effect of learning to operate the powered wheelchair using the sensor systems to assist was demonstrated in the results of these experiments. The negative effect of learning while driving manually and then using sensor systems to support driving was also shown. In the second set of experiments, it was shown that Level 2 support reduced the number of collisions for various obstacle configurations.

Results would have benefited from a larger number of volunteers for the second set of experiments that is ongoing. In addition, results obtained during the second set of experiments are not consistent with some of those obtained during the first set of experiments. In the first set of experiments, it was suggested that support had a positive effect during learning and skill development compared with driving manually. This was not observed in the second set of experiments. Although further experiments focusing on user performance are needed to obtain a more general result, the work presented here has provided an insight concerning the behaviour of powered wheelchair users as they learn under different conditions and with different levels of support provided to them.

It should be noted that although this research has suggested that using the sensor systems during training is efficient, other research has suggested that once a user has become proficient at driving a powered wheelchair then users perform better in unrestricted environments without the sensor systems [23],[27]. The sensor systems become more useful as the environment becomes more complicated.

#### REFERENCES

- [1] D. Chikura, M. Takahashi, S. Watanabe and M. Kitamura, "Adaptation of User Behavior to the Different Level of Tele-Operation Support". *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 3, pp. 739 – 744. 1999.
- [2] T.B. Sheridan, "Telerobotics, Automation and Human Supervisory Control", *Cambridge Massachusetts: The MIT Press*, 1992.
- [3] I.J. Stott, D.A. Sanders, "The use of virtual reality to train powered wheelchair users and test new wheelchair systems". *INT J REHABIL RES*, vol. 23 no. 4, pp. 321-326, 2000.
- [4] A. Kirlik, "Modeling Behavior in Human-Automation Interaction: Why an "Aid Can(and Should) Go Unused," *HUM FACTORS*, 1993, vol.35, no. 2, pp.221-242.
- [5] M. Itoh *et. al.*, "Experimental study of situation-adaptive human-automation collaboration for takeoff safety", *Proc. of 7<sup>th</sup> IFAC/IFIP/IFORS/IEA Symp on Analysis, Design & Evaluation of Man-Machine Systems*, 1998, pp. 371-376.
- [6] I.J. Stott and D.A. Sanders, "New powered wheelchair systems for the rehabilitation of some severely disabled users", *INT J REHABIL RES*, vol. 23, no. 3, pp. 149-153, 2000.
- [7] D.A. Sanders and I.J. Stott, "A new prototype intelligent mobility system to assist powered wheelchair users". *IND ROBOT*, vol. 26, no. 6, pp. 466-475, 1999.
- [8] M.J Goodwin, D.A. Sanders and G. Poland, "Navigational assistance for disabled wheelchair-user". *Proc' Euromicro 95* vol. 43, pp. 73-79, 1997.
- [9] D.A. Sanders, J. Bergasa-Suso, "Inferring Learning Style From the Way Students Interact With a Computer User Interface and the WWW". *IEEE T EDUC*, vol. 53, pp. 613-620, 2010.
- [10] D.A. Sanders, "Controlling the direction of "walkie" type forklifts and pallet jacks on sloping ground". *ASSEMBLY AUTOM*, vol. 28, no. 4, pp. 317-324, 2008.
- [11] D.A. Sanders, I.J. Stott, D. Robinson and D. Ndzi, "Analysis of successes and failures with a tele-operated mobile robot in various modes of operation". *Robotica*, vol. 30, pp. 973-988. 2012.
- [12] P.G Backes, "Supervised Autonomy for Space Robotics," *Progress in Astronautics and Aeronautics*, Vol. 161, 1994, pp. 139- 158.
- [13] J.V. Draper *et.al.* "Measuring Operator Skill and Teleoperator Performance," *Proc. of Int Symp on Teleoperation and Control*, 1998.
- [14] D.A. Sanders, M. Langner M and G.E. Tewkesbury, "Improving wheelchair-driving using a sensor system to control wheelchair-veer and variable-switches as an alternative to digital-switches or joysticks". *IND ROBOT*, vol. 37, no. 2, pp. 157-167, 2010.
- [15] W. Gao and M. Hinders, "Mobile robot sonar backscatter algorithm for automatically distinguishing walls, fences, and hedges". *INT J ROBOT RES*, vol. 25, no. 2, pp. 135-145, 2006.
- [16] D.A. Sanders, I.J. Stott and M.J. Goodwin, "A software algorithm for the intelligent mixing of inputs to a tele-operated vehicle. *J SYST ARCHITECT*, vol. 43 no. 1-5, pp. 67-72. 1997.
- [17] D.A. Sanders, J. Graham-Jones and A. Gegov, "Improving ability of tele-operators to complete progressively more difficult mobile robot paths using simple expert systems and ultrasonic sensors". *IND ROBOT*, vol. 37, no. 5, pp. 431-440. 2010.
- [18] D.A. Sanders, Analysis of the effects of time delays on the teleoperation of a mobile robot in various modes of operation. *IND ROBOT*, vol. 36, no. 6, pp. 570-584. 2009.
- [19] D.A. Sanders *et al.* Simple expert systems to improve an ultrasonic sensor-system for a tele-operated mobile-robot. *SENSOR REV*, vol. 31, no. 3, pp. 246-260. 2011.
- [20] D.A. Sanders, Comparing ability to complete simple tele-operated rescue or maintenance mobile-robot tasks with and without a sensor system. *SENSOR REV*, vol. 30, no. 1, pp. 40-50. 2010.
- [21] G.E. Tewkesbury and [20] D.A. Sanders, "A new simulation based robot command library applied to three robots". *J ROBOTIC SYST*, vol. 16, no. 8, pp. 461-469. 1999.
- [22] D.A. Sanders, "Recognizing shipbuilding parts using artificial neural networks and Fourier descriptor, *P I MECH ENG B-J ENG*, vol. 223, no. 3, pp. 337-342, 2009.
- [23] G.E. Tewkesbury and D. Sanders, "A new robot command library which includes simulation". *J ROBOTIC SYST*, vol. 26, no. 1, pp 39-48. 1999.
- [24] D.A. Sanders, "Comparing speed to complete progressively more difficult mobile robot paths between human tele-operators and humans with sensor-systems to assist". *ASSEMBLY AUTOM*, vol. 29, no. 3, pp. 230-248. 2009.
- [25] J. Bergasa-Suso, D.A. Sanders and G.E. Tewkesbury, "Intelligent browser-based systems to assist Internet users". *IEEE T EDUC*, vol. 48, no. 4, pp. 580-585. 2005.
- [26] D.A. Sanders and A. Baldwin, "X-by-wire technology", *Total Vehicle Technology Conference* pp, 3-12. 2001.

# Serious Game of increase Cognitive Function for Elderly using Arduino based coordinated movement

Joung-Won Ko<sup>1</sup>, Sung-Jun Park<sup>2</sup>

<sup>1</sup>Dept. of Game Engineering, Hoseo University, Asan-si, Chungcheongnam-do, Korea

<sup>2</sup>Dept. of Game Engineering, Hoseo University, Asan-si, Chungcheongnam-do, Korea

**Abstract** -Advanced in the development of medical technology, human life expectancy is increasing and old age related cognitive changes such as dementia and the interest in the improvement of cognitive function is increasing. Because of this, prevention of dementia and efforts to improve cognitive skills associated with dementia is actively walking in technical contents. And also the serious games to prevent dementia are introduced by improving cognitive functions. In this study, in order to compensate the defect of the existing functional game preventing dementia that their effectiveness of dementia is unproven, we have developed the functional game to improve elderly cognitive function based on the coordinate movement using Arduino that improves physical motor function and brain function at once. The effectiveness of this game was tested targeting 20 elderly people in Cheonan community center. In order to verify the effectiveness, the group was divided into the training group with game developed by using touch function of the tablet PC and the training group based on the coordinate movement using Arduino. The effectiveness of the cognitive function improvement, reaction velocity, and increase in the life satisfaction was verified and it was confirmed that the training group based on the coordinate movement using Arduino was more efficient.

**Keywords:** Elderly, Serious Game, Dementia, Cognitive Skills, Arduino, Motion Sensor

## 1 Introduction

As the human life expectancy increases, and the mortality rate decreases due to economic development and improvement of medical technology, population aging is underway. Interest in senile dementia associated with cognitive changes in old age is also increasing because of population aging. The dementia has emerged as a social problem in the modern society where the elderly population gradually increases.

According to the result of analyzing medical fees spent to dementia patients by National Health Insurance Corporation, the number of dementia patients was 105 thousands in 2006 and nearly three times increased to

312 thousands in 2011. It means that the number of patients who underwent dementia treatment increased by 24.4% annually in average.[1] In addition, the number of dementia patients in Korea is forecasted to exceed one million in 2025.[2]

In the past, the elderly with dementia were taken care of at home. However, due to social changes such as individualization and industrialization, as the number of the elderly with dementia who were taken care of at home decreases, and the elderly population continues to grow, the number of the elderly with dementia who need long-term care will rapidly rise as well as the number of the elderly with dementia will steadily increase. Moreover, family members who watch them experience great suffering because of the nature of dementia.[3, 4] Therefore, a dementia prevention program is required to prevent dementia by enhancing judgement, memory, ability to react instantly, and concentration, which are typical cognitive functions associated with dementia, and to improve quality of life of the elderly.[5]

Accordingly, functional dementia prevention games allowing the elderly to receive a dementia prevention training entertainingly and voluntarily by combining a fun element of games have appeared.[6] However, cognitive, physical, and emotional characteristics of human develop not independently but affecting one another, thus to prevent dementia more efficiently, improvement of body's motor skills and brain functions should be made together all the time.[7]

In this paper, a dementia prevention game for the elderly based on the coordination action utilizing Arduino was developed to complement shortcomings of existing dementia prevention functional games whose effectiveness is not verified properly and to improve body's motor skills and brain functions together. In order to verify the effectiveness of the developed game, 20 days of training was conducted targeting the elderly in a welfare center located in Cheonan. The effectiveness of 20 days of training on improvement of judgement, memory, ability to react instantly, and concentration was assessed by comparing the difference between when using Arduino to improve body's motor skills together and when using finger touch only considering enhancement of brain functions only. As a result, it was

found that the effectiveness was superior when using Arduino to improve body's motor skills and brain functions together.

## 2 Related Research

### 2.1 Dementia Prevention Game for the Elderly

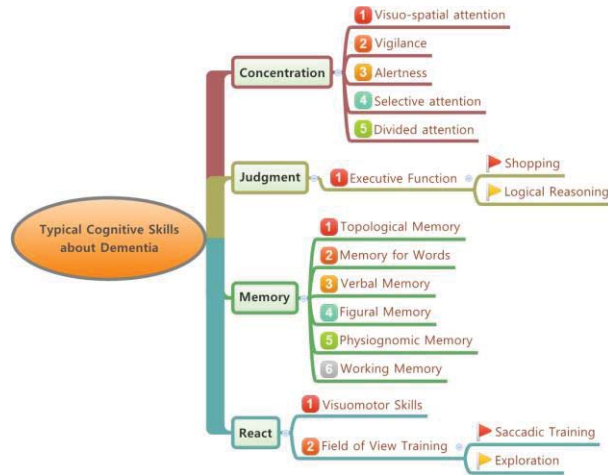


Figure 1: Typical cognitive skills about Dementia

Typical cognitive functions associated with senile dementia include judgement, memory, ability to react instantly, and concentration, thus improving these functions is very helpful to prevent dementia. Dementia prevention games for the elderly contain elements that can improve judgement, memory, ability to react instantly, and concentration. As presented in [figure 1], games should consist of judgement, memory, ability to react instantly, and concentration to assess user's brain functions, and through appropriate training and treatment, dementia should be prevented and improved. [8]

### 2.2 Dementia Prevention Game for the Elderly “Rejuvenating Town”



Figure 2: Dementia Prevention Game for the Elderly “Rejuvenating Town”

The Rejuvenating Town is a dementia prevention game for the elderly containing a total of 9 cognitive games by dividing typical cognitive functions associated with dementia into memory, attention, and judgement. It is developed to train and enhance each element of cognitive functions through visual or auditory signal stimuli along with simple physical activity using touch interface and buttons. The game was developed by advice and thorough verification of medical staff at Asan Medical Center including Min-Ho Geon, division head of rehabilitation medicine, etc. and an expert group at Sungshin Women's University including Kyoung-Chun Lim, professor at geriatric nursing, etc. From July 2012, a clinical trial was performed targeting 100 elderly for 5 months at Asan Medical Center.[9]

### 2.3 Dementia Prevention Game for the Elderly Utilizing Wearable Device

As cognitive, physical, and emotional characteristics of human are not independent but affecting one another, action-based type of games rather than mouse click or touch-based games are more efficiently helpful for dementia prevention.

### 2.4 Arduino

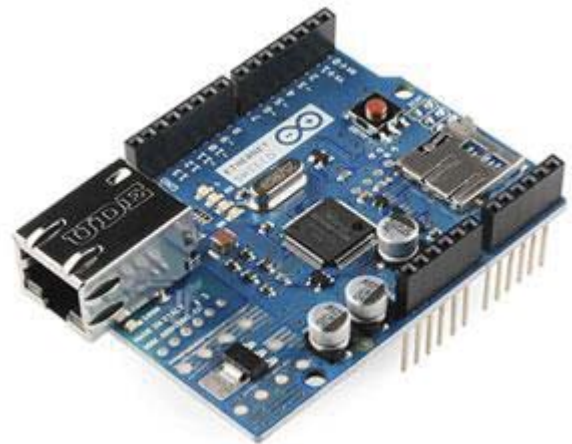


Figure 3: Configuring Arduino basic board

However, the existing wearable device was not made considering elderly users, and therefore, an inclusive design type of wearable device should be produced for dementia prevention games for the elderly.[10] Arduino is AVR-based single board microcontroller, the most suitable open source to make as a type of this inclusive design. In addition, it is providing the integrated development environment (IDE) for software development. Arduino can make a product interacting with the environment by accepting a value from the switch or sensor and controlling external electronic devices such as LED or monitor. Moreover, it can be

used by interlocking with software such as flash, processing, visual programming, Unity3D, etc.[11]

### 3 Dementia Prevention Game for the Elderly Developed in this Study

#### 3.1 Game Design Elements for Dementia Prevention

In this paper, a coordination action based dementia prevention game was suggested as a treatment method for senile dementia prevention. In [Table 1], four brain function elements including concentration, judgment, memory, and ability to react instantly were presented. Additionally, an exercise effect due to muscle use of the elderly can be seen together by playing the dementia prevention game based on the coordination action using Arduino.

Table 1: Four Kind of Brain Skill elements for Dementia

Brain Function	Competence Test	Daily Training
Concentration	Concentration Test	Dugg'y Hide-and-Seek
Judgment	Judgment Test	Nyang's Separate
Memory	Memory Test	Guri's Twinkle Twinkle
React	React Test	Grandpa Pile Tower

#### 3.2 Arduino Hardware Design Part



Figure 4: Arduino Bracelet and Anklet

To make an action recognition ring and bracelet using Arduino, as shown in [figure 4], for the bracelet, mpu-6050 acceleration sensor and bluetooth FB155BC were installed in Arduino pro mini board, and for the ring, mpu-6050 acceleration sensor was installed in Arduino beetle.

#### 3.3 Arduino Software Design Part

As shown in [figure 5], Arduino can do code work to make a desirable program using the exclusive compiler of Arduino. In this study, codes for a ring worn on the finger and a bracelet worn on the wrist were written respectively. For the bracelet, codes for actions including left and right rotation below the wrist, left and right rotation of the wrist, and up/down action of the wrist were written.

```

Gamedevice | 아두이노 1.0.6
파일 편집 스케치 도구 도움말

Gamedevice i2c kalman.h
// atan2 outputs the value of -π to π (radians) - see http://en.wikipedia.org/wiki/
// We then convert it to 0 to 2π and then from radians to degrees
accXangle = (atan2(accY,accZ)+PI)+RAD_TO_DEG;
accYangle = (atan2(accX,accZ)+PI)+RAD_TO_DEG;

double gyroXrate = (double)gyroX/131.0;
double gyroYrate = -((double)gyroY/131.0);
gyroXangle += gyroXrate*((double)(micros()-timer)/1000000); // Calculate gyro angle wi
gyroYangle += gyroYrate*((double)(micros()-timer)/1000000);

kalAngleX = kalmanX.getAngle(accXangle, gyroXrate, (double)(micros()-timer)/1000000);
kalAngleY = kalmanY.getAngle(accYangle, gyroYrate, (double)(micros()-timer)/1000000);
timer = micros();

int yyy = kalAngleY; // Roll
int xxx = kalAngleX; // Pitch
    
```

Figure 5: Arduino Gesture Recognition Code

#### 3.4 Interlocking between Arduino and Unity

```

ArduinoManager.cs CameraMove.cs
selection
22 public void OpenPort(string portName, int baudRate)
23 {
24     if ( serialPort == null )
25         serialPort = new SerialPort();
26
27     serialPort.PortName = portName;
28     serialPort.BaudRate = baudRate;
29     serialPort.DataBits = (int)8;
30     serialPort.Parity = Parity.None;
31     serialPort.StopBits = StopBits.One;
32     serialPort.Open();
33 }
34 public void ClosePort()
35 {
36     if ( serialPort == null || !serialPort.IsOpen ) return;
37
38     serialPort.Close();
39     serialPort.Dispose();
40     serialPort = null;
41 }
42 public void Send(string message)
43 {
    
```

Figure 6: Arduino interlock with Unity3D

The dementia prevention game suggested in this paper was created based on Unity3D. To make the action recognition device, Arduino, recognized in the game, it receives communication information that is sent from

Arduino device within Unity3D game and applies it into the game through operation.

### 3.5 Dementia Prevention Game for the Elderly, “Youth over Flowers”

A dementia prevention game for the elderly, named Youth over Flowers, was developed combining four elements to prevent dementia. The [figure 7] displays the overall structure of Youth over Flowers in diagram form.

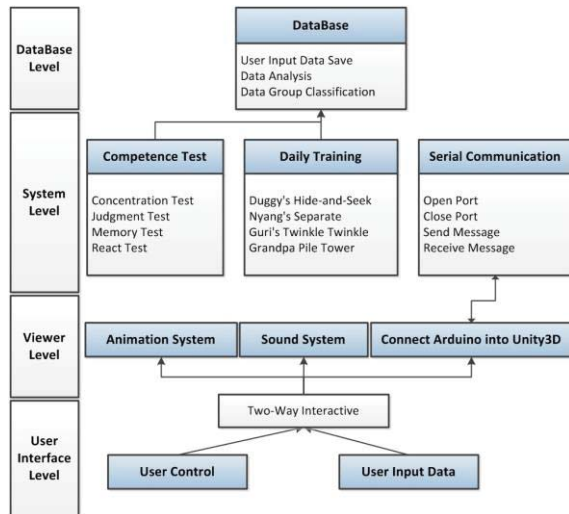


Figure 7: “Youth over Flowers” System Architecture

The game Youth over Flowers is divided into proficiency testing phase and every day training phase. In the proficiency testing phase, the user's levels of concentration, judgement, memory, and ability to react instantly are evaluated, while in the everyday training phase, the user continues to play for 20 days to improve concentration, judgement, memory, and ability to react instantly. The [figure 8] shows the main screening of Youth over Flowers, the functional game for elderly cognitive function improvement.



Figure 8: “Youth over Flowers” Main Screen

As shown in [figure 9], the judgement test game is finding the same color and the same picture with the three pictures coming out on the top and designed to

measure judgement not only putting drawings but also having background color.

In the memory test game, a picture is divided into four pieces, and one piece is shown at a time. The user finds the same picture with the whole picture in that all pieces are together, and it is designed to measure memory.

The test for the ability to react instantly is a game shooting a slingshot when the passing target is located in the center and is designed to measure the ability to react instantly.

In the concentration test game, moles randomly come up from 9 holes, and the user hits the moles, which is designed to measure concentration.



Figure 9: Touch-based Skills test game

The game utilizing Arduino based on the coordination action is designed to select an object that the user wants. In case of the bracelet, using actions moving left and right, an object that the user wants can be selected, while in case of anklet, actions moving into eight directions are used.

As presented in [figure 10], to enhance cognitive functions more efficiently by connecting cognitive, physical, and emotional characteristics in the process of changing touch-based game into coordination-action based game, the graphic design was newly constructed to match the emotion of the elderly.



Figure 10: Arduino-based training game

If four tests, judgement, memory, ability to react instantly, and concentration, are completed, as shown in [figure 11] left, comprehensive information about the test is displayed, and based on the data in database, the information about whether it is lower or higher than the average can be checked. Moreover, after determining which part should be focused on in training based on the test result, it constructs a training game as shown in the [figure 11] right.



Figure 11: Test result(left), Program construction(right)

In this paper, the efficacy of Youth over Flowers, a dementia prevention game, implemented in this study, was verified. For this, 3-month experiment was conducted targeting 20 senior citizens in a welfare center located in Cheonan.

For the dementia prevention game, four kinds of cognitive function test games and four training games to improve cognitive functions were arranged. In this experiment, for group 1, 20 days of training was conducted using tablet PC touch only, and for group 2, it was conducted based on the coordination action interlocking with Arduino.[figure 13]



Figure 13: Play the training game in welfare center

## 4 Experiment and Result

	Content			
<b>Period</b>	June to August 2014			
<b>Place</b>	Welfare center in Cheonan			
<b>Target</b>	Welfare center 65-year-old or older(20 people)			
	Age	Male N %	Female N %	Total N %
	60	5 25	6 30	11 55
	70	7 35	2 10	9 45
	Total	12 60	8 40	20 100
<b>Sub-stance</b>	Efficacy verified of Dementia Prevention Game for the Elderly			
<b>Experimentation</b>	For the dementia prevention game, four kinds of cognitive function test games and four training games to improve cognitive functions were arranged. In this experiment, for group 1, 20 days of training was conducted using tablet PC touch only, and for group 2, it was conducted based on the coordination action interlocking with Arduino.			
<b>Method</b>	Improvement of Cognitive Functions, Measurement of Response Time and Accuracy, Emotional Depression of the Elderly  Group 1 : Touch-based Group 2 : Arduino-based			

Figure 12: Experiments overview

### 4.1 Difference in Improvement of Cognitive Functions between Touch-based and Action Recognition-based Game

The differences in cognitive functions between the initial test game and after completion of 20 days of training were compared between the group 1, which used touch-based small muscles only, and the group 2, which received a training based on the coordination action using Arduino.

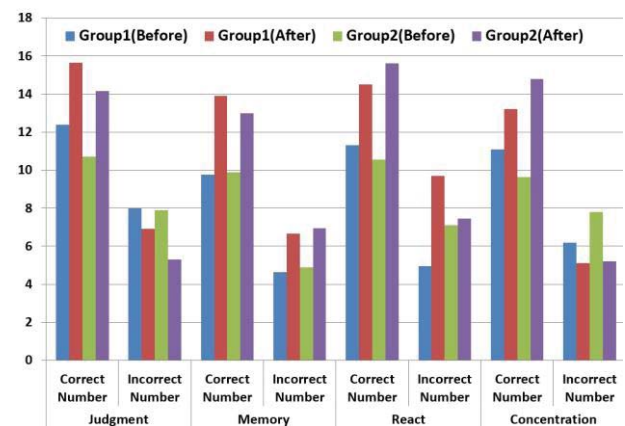


Figure 14: Improvement of cognitive functions between group 1 and 2

## 4.2 Measurement of Response Time and Accuracy

The differences in response time and accuracy between the initial test game and after completion of 20 days of training were compared between the group 1, which used touch-based small muscles only, and the group 2, which received a training based on the coordination action using Arduino.

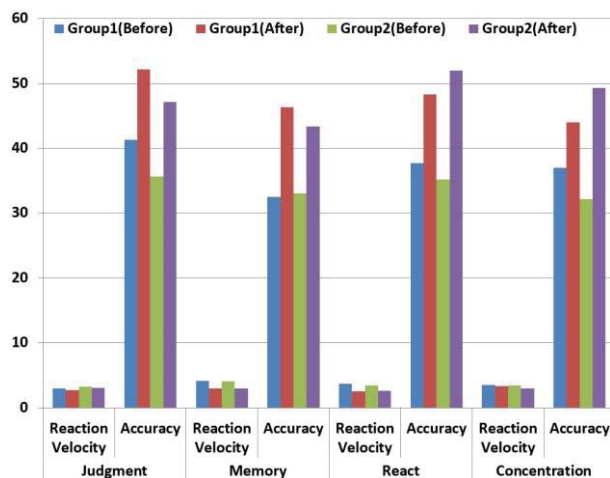


Figure 15: Measurement of Response Time and Accuracy between group 1 and 2

## 4.3 Impact on Emotional Depression of the Elderly

To investigate the difference in the change of life satisfaction between using touch-based small muscles and conducting the training based on the coordination action utilizing Arduino, the degree of depression was tested for group1 and group 2 at initial test game and after completion of 20 days of training.

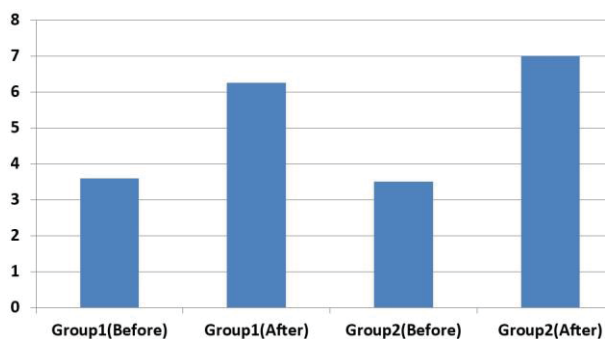


Figure 16: Emotional Depression of the Elderly between group 1 and 2

## 5 Conclusions

In this study, a dementia prevention game for the elderly was developed combining four kinds of cognitive function improvement factors with a fun factor of the game in order to prevent dementia by enhancing typical cognitive functions associated with dementia, judgment, memory, ability to react instantly, and concentration, and by improving quality of life of the elderly. Additionally, a wearable device utilizing Arduino, as a method improving brain functions and body's motor skills at the same time, was made and incorporated into the game because cognitive, physical, and emotional characteristics of human develop not independently but affecting one another.

To verify the effectiveness of the game developed, 20 days of training was conducted targeting 20 senior citizens at a welfare center located in Choeran, dividing into group 1 using tablet PC touch function only and group 2 receiving a training based on the coordination action interlocking with Arduino. Through this experiment, it was confirmed that the degree of improvement of judgment, memory, ability to react instantly, and concentration was greater when the coordination action based training was given using Arduino than when small muscles were used by simple touch. As a result of measuring the degree of depression at initiation of the training and after completion of 20 days of training, it was found that the quality of life increased more when using Arduino.

In the future, further study will be performed to additionally develop a game that can improve cognitive functions and to make it more helpful in preventing dementia and keeping physical health by adding more efficient actions for the physical exercise effect through studies of actions of the elderly.

## 6 References

- [1] Yong-Chul Kwon, Jong-Hwan Park, "Standardization of Korean of the Mini-Mental State Examination (MMSE-K) for Use in the Elderly. Part I. Development of MMSE-K"; Korean NeuroPsychiatric Association, 28(1), 125-135 (1989)
- [2] Ministry of Health & Welfare. National dementia strategy 2013-2015[Internet]. Seoul: Author; 2012 [cited 2013 May 10]. Available from: [http://www.mw.go.kr/front\\_new/al/sal0301vw.jsp?PAR\\_MENU\\_ID=04&MENU\\_ID=0403&BOARD\\_ID=140&BOARD\\_FLAG=00&CONT\\_SEQ=274723&page=1](http://www.mw.go.kr/front_new/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&BOARD_ID=140&BOARD_FLAG=00&CONT_SEQ=274723&page=1).
- [3] Lee YW, Park KH, Seong YS. "A study on changes of primary caregivers' fatigue, depression and life satisfaction by using dementia day care services", Journal of Korean Academy of Adult Nursing, 20(3), 443-451 (2008)
- [4] Kwon JD, Go HJ, Lim SE, Lee SH, Chang WS, Lee YJ. "Dementia and family", Seoul: Hakjisa Publisher, (2002)
- [5] Yun-Hee Son, "(A)Study on the effect of the dementia prevention program for senior citizens", Graduate School of Chosun University, Master's degree paper (2013)
- [6] Kyu-Man Jeong, "Prevention of Digital Dementia using a Serious Game", Journal of Korean Society for Computer Game, 26(4), 153-157 (Dec 2013)
- [7] Suk-Hee Jang, "(A)Study on the effect of the dementia prevention program", Graduate School of Chosun University, Master's degree paper (2007)
- [8] Jin-Ho Kim, "(The)Study on effectiveness of game method for dotard", Graduate School of Industry Information Mokwon University, Master's degree paper (2005)
- [9] Ji-won Lee, Sung-Jun Park, Nam-Hyun Jo, Kyung-Sik Kim, "Effective Serious Game for Elderly design method", Korea Game Society, Journal of Korea Game Society spring conference (2013)
- [10] Jong-Sun Lee, Jae-Beom Park, Jun-Dong Jo, "A Study on Design of Wearable Healthcare Device for Elders who Live Alone focused on Inclusive Design", Journal of Korean Society of Design Science, 2014(5), 176-177 (May 2014)
- [11] Chae-rim Park, Sang-Jin Lee, "Interaction light design with Arduino: Light design for Interaction elements", Journal of Korea HCI Association, 2014(2), 963-965 (Feb 2014)



# Autonomous Wheelchair Navigation Prototype with an Arduino Robot

Anna Shafer, Michael Turney, Francisco Ruiz, Justin Mabon, Michael Nooner, Yu Sun, Vamsi Paruchuri

Dept. of Computer Science, University of Central Arkansas, Conway, AR 72034, USA

**Abstract** - Navigating through a large and complicated hospital can be difficult to most people, especially to those elderly and/or disabled patients. To help patients more efficiently while reducing the manpower, in this research, we have proposed and developed an Autonomous Wheelchair Navigation Prototype with an Arduino robot for hospital navigation. With a user-friendly interface, the proposed prototype is able to determine the optimal path to find locations accurately and can successfully control robots' movement during the navigation. Thus, it can remove the need to learn the ins and outs of hospitals and improve the quality of life for its users. Our prototype for this system has shown good preliminary results and is looking towards a bright future.

**Keywords:** Arduino, Robot, Wheelchair, Autonomous, Navigation

## 1. Introduction

### 1.1 Background

The United States Census Bureau released the status of people with disabilities in July of 2012. Over 3.5 million people use a wheelchair to assist with mobility throughout their daily lives. Depending upon the severity and type of disability, mobility and independence can be a challenge. This issue is compounded by the aging population of the United States. The post World II baby boom of 1946-1964 flooded the United States alone with 75 million births. The Baby Boomers are beginning to hit the 65 and older mark which will cause an influx of patients being admitted into hospitals. As such health care resources in the US are not prepared to meet such a rapid increase in elderly patients. These patients will require more and more medical care. The situation is already taking a toll on health care resources, requiring more health care personnel to accommodate the rising number of elderly patients [1].

The draw on human resources in the US as a result of this population imbalance is uneconomical. In comparison, Japan is ageing faster than any other country in history, with vast consequences for its economy and society [2]. They have, however, proposed a solution to some of the issues that we are now beginning to face. One such problem is the navigation of hospitals by elderly patients.

Navigating large hospitals can be difficult for patients and arduous for the caretakers of elderly patients. Without assistance, hospital navigation can be difficult or impossible for them. The Autonomous Wheelchair Prototype is the solution to this problem.

### 1.2 Related Work

Today's technological advances have opened many doors for those with physical impairments. However, current technologies are still lacking. Manual wheelchairs are still the standard in hospital settings, requiring a patient to be in good physical condition and have knowledge of the layout in order to navigate or have assistance from health care personnel. Several solutions to this problem have been proposed.

#### 1.2.1 MICA

The Mobile Internet Connected Assistant (MICA), developed by the Lulea University of Technology in Sweden, allows users to operate a wheelchair with movement of the head, voice commands, or fully autonomously [3]. It is designed to be controlled remotely, over the internet, or by the user himself. Where the MICA suffers though, is its lack of pathfinding. The only autonomous navigation the MICA is capable of voice recognition which requires a user to dictate very specific commands. A user is required to be able to speak or make head movements to navigate this device and to have a knowledge of the environment in order to direct it.

#### 1.2.2 RobChair

The RobChair navigation system was developed at the University of Coimbra, Portugal. The system is designed to assist quadriplegic or simultaneously blind and paraplegic people with their mobility and navigation in domestic environments [4]. The system is voice activated and is equipped with obstacle avoidance to allow for general navigation commands to *go right* or *go left*. However, this system lacks any knowledge about its surroundings. It requires the user to have a thorough knowledge of their environment and be able to navigate to their destination.

#### 1.2.3 Aviator

The Aviator is a wheelchair designed by Hung Nguyen and his team at the Centre for Health Technologies at the University of Technology, Sydney. This hands-free

wheelchair uses an electroencephalography (EEG) to read and translate brain signals into navigational commands for the wheelchair [5]. While the wheelchair can navigate by thoughts alone and uses cameras to avoid obstacles, it is very expensive and made only for patients who are severely physically handicapped. It also requires the user to have a prior knowledge of their environment.

### 1.3 Motivation and Solution

Hospitals are huge and difficult to navigate. The long hallways and labyrinthine passageways are problematic for patients regardless of how many times they may have visited. Patients often have to rely on others to help them navigate, either by physically pushing them in a wheelchair or by guiding them to their destination.

Our proposed autonomous wheelchair navigation system will be able to transport a patient to their destination with only the push of a few buttons. This system will use RFID tags strategically embedded throughout the hospital to provide orientation information to the wheelchair device. Our system takes advantage of the static nature of hospitals by allowing previously populated floor maps of hospitals to be downloaded to the device at any time, allowing the device to navigate to any destination choice without requiring prior knowledge of the environment's layout from the user. RFID tags will be used because they are cheap, readily available, and will provide more accurate feedback than a GPS in this indoor setting.

Any patient will be able to use this system. It is designed to remove the need of learning hospital layouts. By allowing technology to assist in this way, better management of human and monetary resources will be possible. This wheelchair navigation system will save money for health care administrations while accommodating the needs of a wider range of patients than is possible with only a manual wheelchair.

The rest of this paper is organized as follows. Section 2 describes the overall system structure. Section 3 details the prototype's components. Section 4 discusses the initial results and Section 5 concludes the paper.

## 2. System Overview and Description

Fig. 1 illustrates the system overview. It starts with the Administrative UI (AUI), where a hospital administrator can create maps of a hospital. He can build maps to mimic floor plans and save them to an online repository, which contains all the floors and buildings of a hospital along with a database that links room locations and room numbers together. The database also links RFID values and tag numbers, so the admin only has to memorize tag numbers but not the 12 digit hexadecimal RFID values.

When a user sits in their wheelchair, the map of the floor is downloaded to the device. When the user inputs the room number they wish to go to, the wheelchair will then

scan their current location and generate the optimal path to their destination. It will then send directions to the wheelchair's motors directing the wheelchair. When the wheelchair drives over an RFID card, it will compare the value with the virtual map. If the RFID card matches then the wheelchair is in the correct location and is given another direction. If the values don't match the wheelchair will update the path with the new location. Once the wheelchair reads the RFID card of its destination the chair will stop.

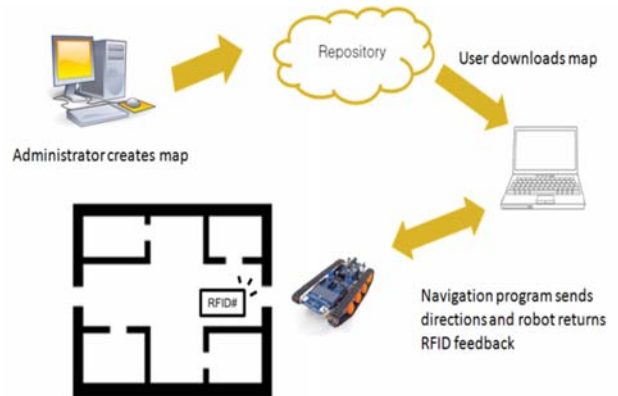


Fig. 1: Diagram of robotic wheelchair simulation

## 3. Proposed Prototype Components

The prototype wheelchair navigation system is broken up into four modules (Fig.2): the Administrative UI, the Software Keypad, the Navigational algorithm, and the Robot. The Administrative UI is used to create the maps of hospital floors. The Software Keypad is used to retrieve the room locations from the database. The location is sent to the Navigational Algorithm which determines the optimal path. The algorithm sends instructions to the Robot. The Robot moves according to the instructions and sends feedback to the algorithm to update its position.

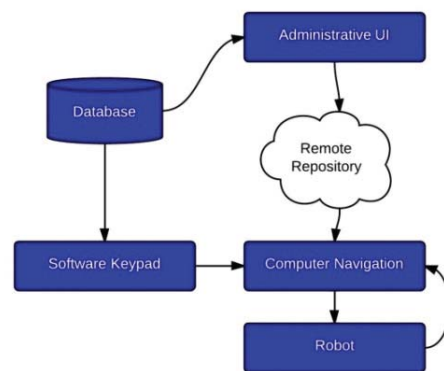


Fig. 2: Prototype wheelchair system flow

### 3.1 Administrative UI

The Administrative UI is designed in three stages: generate, build, and save. The administrator must first generate a map of a set size. Once the map is generated, the

administrator can then start to build the map. By using a toolbar of pre-defined shapes, the admin can reconstruct most floor plans. The admin is able to click the shape in the toolbar and place it in the map space. Doors and RFID-embedded tiles require the admin to input an RFID tag, which is a simple number assigned to each tag. During and after the building stage, the admin can save their work. The admin can then load the map at any time to continue work. When the map is finished and saved, it will then be in the online repository and it is ready to be used.

The AUI (Fig. 3) is created using JavaScript, PHP and HTML. The HTML is used to create a canvas for the map and toolbar to be drawn in. The JavaScript populates the canvases with the map and toolbar. The JavaScript is then used to draw the map by the admin clicking the tile in the toolbar and then clicking a location in the map. The load function uses JavaScript to open up a text file and parse the contents to find the value of each position in the map. The PHP is used to save the map. It also converts the RFID tag the admin inputted into the 12 digit hexadecimal RFID code. The PHP also finds the rooms in the map and stores the location of the room in a database. This database is used later by the software keypad to retrieve end-point locations.

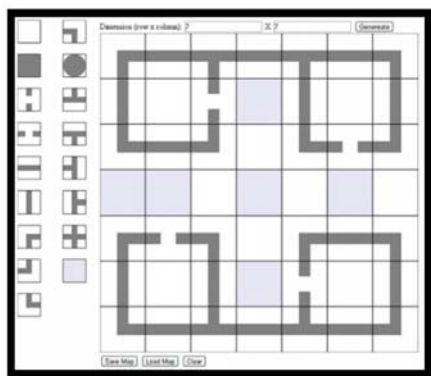


Fig. 3: GUI example map

From the Administrative UI Web Page, the admin can generate a map and start drawing. Fig. 4 shows the three main functions and how they are called. The Generate function will create an empty map for the admin to draw in. At this point the admin can then click around on the web page. Depending on where the admin clicks will determine which function is called. If the admin clicks in the toolbar, selectShape() is called to load a function into a variable. If the admin clicks inside the map space, drawShape() will be called.

The UI contains an array of functions. These functions are used to draw the many different shapes. When the user clicks in the toolbar, that function from the array is loaded into a variable. When the user clicks inside the map, the function in the variable is called. This lets the user draw many shapes in the map without having a large if-else chain to slow it down. The map is a 2d array that contains strings. The value of the string can range from "0" to "16" which represent the possible shapes. Values of "2", "3", and "16"

all contain extra information. "2" and "3" are doors and must contain the RFID value in the floor tile and the room number. "16" is the RFID embedded tile and as such contains the RFID value. An example of a door tile is "3;2;101" which means a door to room 101 has the RFID tag 2. An RFID embedded tile would look like "16;4" which means the tile has the RFID tag 4.

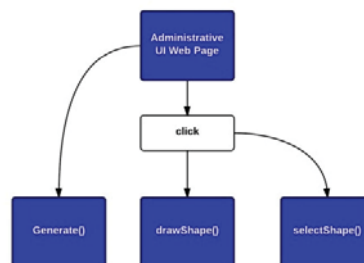


Fig. 4: Three Main Functions

When the user is done and wants to save, the save function is called. This function sends the 2D array representing the map to a php script. This script will go through the array and write to a text file the contents. It will place the location of the tile and its value. A value of 'w' means wall, a value of 'e' means empty, no value represents an empty space, and anything else is the 12 digit hexadecimal value. When the program detects there will be a 12 digit hexadecimal value, it will connect to a database that pairs the hexadecimal value with the RFID tag number. This allows the user to not have to memorize so much. An example of an entry in the text file is "2 2 45DCA0112385" which means that position 2,2 has that RFID value while "2 3 w" means there is a wall at 2,3. When the text file is completed, the user can then operate the software keypad to find room locations.

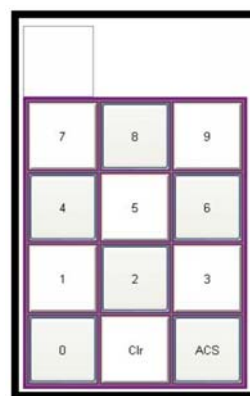


Fig. 5: Software Keypad example

### 3.2 Software Keypad

Our keypad (Fig.5), programmed in HTML and Javascript, contains mostly numbers and has twelve buttons, which is easy for users. When a patient uses our Autonomous Wheelchair Prototype, they would press on the keypad the room number they wish to go to. The keypad sends the room number to the remote map server. That room

number is looked up in the database and the database returns the location and saves that location into a text file. The robot then navigates to the destination that the patient wanted to go to.

### 3.3 A star Algorithm

The A star (A\*) algorithm is standard in navigation. It finds the optimal path between two points in a known map which consist of points or nodes. These nodes are spread out in the map and serve as start and end points as well as all points in between. Each node also contains priority and level values. The priority value is equal to the distance it is from the end point added to its level. The level of a node is how far away the node is from the starting point. So a priority of 15 means that the node is 15 units away from the end point, while a level of 11 means this node is 11 units away from the start point.

The standard A\* algorithm is not suitable for our project, so we had to make some revisions. One of those revisions is on the node structure. The node structure used in our A\* algorithm contains six properties. These properties are: the X and Y coordinates of the node; the level and priority of the node; the RFID value of the node, and the direction that the node is pointing in. The direction is a value between zero and seven. Zero represents map east and each increment represents another direction in a clockwise direction such that seven will represent map north-east (see Fig. 6). This value is updated to show where the next node in the path is located. The major steps of the revised A\* algorithm are: (1) Begin with the starting Node; (2) Search each neighbor node of the selected node. Start to the east of the node and continue clockwise until all eight directions are checked; (3) For each neighboring node, check to see if they are a wall node or have been visited before; (4) If the node is not a wall node and has not been visited before, place the node in a possible path queue; (5) If the node in the queue has been considered before, compare the two priorities of the paths; (6) Update the nodes priority and direction based on the lowest priority; (7) Place the nodes in a stack; (8) Select a node from the stack and go back to step 2; (9) Once the stack is empty or the end node is found backtrack using the direction on the nodes; (10) This will give the correct path.

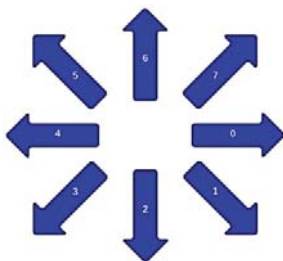


Fig. 6: Eight directions & numbers

The text file from the Administrative UI is required for the A\* algorithm. Before the path can be generated, the program has to load the map into memory. It will parse

through the text file and generate nodes based on the map information. When the path is found, the algorithm will return a string of directions. It will then send one direction at a time to the robot. These directions will indicate which way the robot should be heading. It follows the same numbering pattern from before where zero is map east. When a direction is sent to the robot it will wait for the robot to send back a 12 digit hexadecimal value. The program will take this value and check to see if the robot is in the correct spot. If it is, the program will send the next instruction. If the robot is off course, the program will update the path with the new starting location and send instructions to the robot.

### 3.4 Hardware

The hardware platform used in this initial prototype is based on the DFRobotshop Rover shown in Fig. 7. This robotic kit is constructed around an Arduino Uno microprocessor and its printed circuit board (PCB) which supports the control of the robot's motors, sensors, input and output ports, as well as communication with an external computer via mini USB. Our application takes advantage of the motor controller electronics as well as multiple serial lines. The Arduino microprocessor can be programmed using open source Arduino libraries adapted with C++.

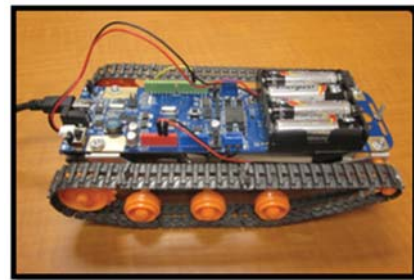


Fig.7: Assembled DFRobotshop Rover

The embedded motor controller chip on the PCB board is used to send alternating signals to the motors corresponding to the desired speed and direction. The robot is directed using a relative positioning algorithm that translates directional navigation commands from the computer into right and left turns relative to the robot's current position. After the command is translated, it is amplified by a factor determined by the robot's terrain and other variables such as the currently supplied power.

The prototype also utilizes a radio-frequency identification (RFID) reader (Fig. 8) as position feedback for the navigational algorithm. This RFID reader operates on a 125 kHz frequency allowing it to read standard electromagnetic card tags (Fig. 9). All communication with the RFID reader is performed through a serial line connected to the microprocessor. The tag numbers are read from the serial stream when they are detected. The tags consist of 16 total bytes in the format shown below. The RFID reader is connected to the PCB as shown in Fig. 10.

*[start of text] – [12 bytes of hex] – [new line] – [carriage return] – [end of text]*



Fig.8: RFID reader

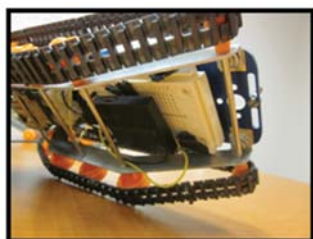


Fig.9: RFID reader on robot

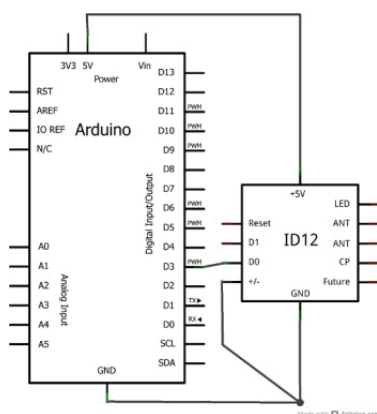


Fig.10: Wiring between RFID reader & Arduino

The robot is only capable of communicating through one serial line at a time. In addition, the USB cable is the communication line for navigational commands to be sent to the robot and the RFID numbers to be sent back to the computer. Our program uses the Windows API to open a COM (communication) port on the computer to receive the information from the microprocessor. All sending and receiving serial pairs must use the same baud rate throughout the communication process. The serial stream read by the program is error prone; therefore, it is necessary to add error correction to the serial processing. We implemented a basic error correction into our prototype; however, it is still possible for enough data to be lost that the only recovery option is an error message.

Fig. 11 illustrates the communication flow throughout the program. A communication handshake is performed at the beginning to verify that the synchronization has taken

place. It lights up an LED to confirm that the handshake was successful and the robot is ready to begin communication with the computer. Then, RFID numbers are sent to the computer for position feedback and they receive direction commands in return until the destination is reached.

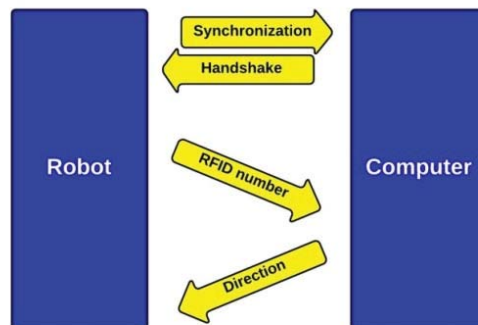


Fig. 11: Communication between robot & computer

### 4. Experimental Results

We tested our prototype of the autonomous wheelchair system according to two metrics. In order for the system to be successful, it must be able to quickly and accurately deliver the patients to their destinations. Thus, the navigational algorithm must be able to generate the path in a timely manner, and the robot must be able to reach the destination accurately 100% of the time.

We performed a worst case scenario test to determine whether the navigational implementation was timely. We created a random map with 5600 nodes. Considering hospitals can be as large as 500,000 square feet and can contain up to ten floors. Each floor would be around 50,000 square feet. Since a single node can represent a doorway which is about three feet standard. This would bring a representation of a hospital floor to a 75 nodes by 75 nodes map. Based on estimations of real world applications, RFID tags accounted for 24% of those nodes and 25% of them were designated as walls. The algorithm to generate the path was timed by subtracting the total milliseconds at the time that it started calculating the path from the total milliseconds that had occurred when the path was done generating. The path was generated, on average, in 840 milliseconds. The average human reaction time to change is around 240 milliseconds. So very shortly after the average human would react to entering their destination, the chair will already have the path generated.

Since the map generated is randomized, it is disorganized and contains many more possible paths than in reality. There are very few intersections in hospitals where there are eight possible ways to go, but since the map is randomized this situation shows up much more often. In a map of a real hospital with optimized RFID placement, the time to traverse the map and find the optimal path will be greatly lowered. Even in the worst case, it still takes less

than one second for the path to be generated and the wheelchair to start moving. When tested on more suitable rooms for the prototype, the path finding time was greatly reduced. For a 7x7 room the path was found in less than 16 milliseconds every run.

In addition to speed, accuracy of the path that was generated and the ability of the robot to follow that path was measured. Measuring how often the robot was able to correctly navigate to the destination was difficult to quantify due to the imprecise nature of the hardware used. To account for these dependencies in our testing, we divided the measurement of the correctness of the navigational algorithm's path generation from the actual hardware's response when designing our tests.

We tested our prototype three times on each of three different maps. The maps were designed to test the ability of the algorithm to correctly find the most efficient path in several different situations and the ability of the RFID reader to provide accurate positioning feedback. For each of the nine test runs, the number of correctly read RFID tags was divided by the total number of tags to be read for the generated path. These ratios are listed in Table 1.

Table 1: Ratio of correctly read to total RFID tags

	Test 1	Test 2	Test 3
Test Map 1 (4 rooms)	0.64	1.00	0.86
Test Map 2 (2 rooms)	1.00	0.89	0.88
Test Map 3 (1 room)	1.00	1.00	0.80

The average for each test map was then found. The average RFID tag reading accuracy for test maps one, two, and three was 0.800, 0.913, 0.923, respectively. The robot's overall RFID tag reading accuracy was calculated by taking the average of these three numbers, yielding an accuracy rate of 87.9%.

The results gathered from our tests so far, support further investigation into this wheelchair navigation system. On nine different tests, requiring the navigational algorithm to calculate different paths, the average run time was less than 16 ms. The time to generate a path in a worst case scenario was still less than a second. In the test cases, the accuracy of the robot was limited only by the hardware. The robot successfully navigated to its destination every time, unless the hardware failed, illustrating the potential of the system.

## 5. Conclusion

Navigating hospitals is a tedious and demanding task for patients and caretakers. With an increasing number of people aging, there will be a greater demand for health care resources in the near future. To alleviate the burden on

health care, an autonomous wheelchair navigation system is very crucial.

We have investigated and developed a prototype for this system. Our Autonomous Wheelchair Navigation prototype was a success. It is able to navigate to an end point following the optimal path. The prototype was built with cheap, off the shelf components. In addition, it is able to generate an optimal path to its destination, which can autonomously traverse to that destination. The prototype successfully simulates the entire system by supplying the platform for an administrator to generate floor maps and save them to an online repository, a user download those maps to the prototype, and letting the robot travel through the map.

The core functionality of prototype is in place, but there are improvements to be made in the future. Bundling floors together in the administrative UI will reduce load times between multi-floor navigation. Adding additional sensors to the robot will help with turning and timing issues along with collision detection. Our initial prototype simulates an autonomous wheelchair hospital navigation system and demonstrates the great potential such a system has to assist our growing elderly population.

## 6. Acknowledgements

This research was partially supported by NSF Award# 1062838: "REU Site: HIT@UCA: Applied Research in Health Information Technology."

## 7. References

- [1] "Chance of Becoming Disabled - Council for Disability Awareness", [http://www.disabilitycanhappen.org/chance\\_s\\_disability/disability\\_stats.asp](http://www.disabilitycanhappen.org/chance_s_disability/disability_stats.asp).
- [2] "Into the unknown | The Economist", <http://www.economist.com/node/17492860>
- [3] M. Hanlon, "The autonomous wheelchair raises the promise of assistive mobile robots," *Gizmag New & Emerging Tech. News*. <http://www.gizmag.com/go/6626/>, Dec. 16, 2006.
- [4] G. Pires, N. Honório, C. Lopes, U. Nunes, A. T Almeida, "Autonomous Wheelchair for Disabled People", *Proc. of IEEE International Symposium on Industrial Electronics*, Guimarães, Portugal, July 7-11, 1997, pp. 797-801.
- [5] PRETZ, K. (n.d.). Building Smarter Wheelchairs: Making life a little easier for people who can't walk. *The Institute*. <http://theinstitute.ieee.org/technology-focus/technology-topic/building-smarter-wheelchairs>

# Perception of motion-lag compared with actual phase-lag for a powered wheelchair system

D.A. Sanders *FIET FIMechE FHEA*, M.Langner *MIET*, A. Gegov, *MIEEE*

**Abstract** — As analogue powered wheelchair systems are digitized and sensors and other systems are added then a lag can be introduced into the control of the wheelchair. This paper discusses phase lag in a powered wheelchair system and when a wheelchair driver might begin to perceive a lag in the motion of the wheelchair. The reduction of any time lag is important for driving performance and ability. The threshold of allowable time lag for a wheelchair driver has not been explored and this work investigated the minimum time lag before a wheelchair driver perceived a lag.

## I. INTRODUCTION

Performance is important when driving a powered wheelchair. Time lag between the input (usually a joystick) and desired output (usually motor control) is one of the factors influencing driving performance. If the driver is able to perceive a time lag then their driving performance is affected. A threshold value of perceived time lag exists  $t_{\min\text{lag}}$ , after which a user realizes that a lag exists. Perception may also depend on phase lag as well as time lag because as an input frequency increases, system phase lag may increase. So, a driver may perceive a motion lag even if the system only has a relatively small time lag. In this research, the threshold value of the time lag is  $t_{\min\text{lag}}$ , the “perceived lag time”. The threshold value of the phase lag is  $\phi_{\min\text{lag}}$  “perceived lag phase”. Time lag in a driving system has been studied by Chang [1], Kawamura [2] and Sanders [3]. But, perceived lag time has only rarely been considered, an exception being Toyoda *et al* [4] who considered motion lag in a tele-operated robot system. It has never been considered for a powered wheelchair. In addition, little attention has been given to the perceived effect of phase lag [4].

## II. THE POWERED WHEELCHAIR SYSTEM

The apparatus consisted of a dedicated controller with analogue interfacing, DC servo-amplifiers and joystick, and a BobCat II powered wheelchair modified to include extra control and sensor systems. Two driven wheels were at the front and two trailing castors at the back. Ultrasonic sensor pairs were mounted over each driving wheel. Altering the differential of

rotational speed of the driving wheels affected steering and direction of movement.

Sonar sensors have been widely used for powered-wheelchairs and mobile powered wheelchairs [5],[6] and ultrasonic ranging was selected, as it was simple, cost effective and robust. Ultrasonic transmitter and receiver pairs were mounted at the front of the powered-wheelchair. With suitable processing the ultrasonic signals were converted to a simple representation of the environment ahead of the wheelchair. An integral function was used with the joystick signals so that the tendency to turn when approaching an object could be overruled by the user, for example to reach a light switch on a wall.

Software algorithms to intelligently mix the inputs to the powered wheelchair (joystick and sensors) were described in [7]–[11] and the wheelchair was driven under computer control by “fly-by-wire”. The direct link between the powered wheelchair and joystick was severed and a computer processed control information. Sensors were activated and interrogated by the computer and the computer was programmed to modify the powered-wheelchair path. Alternatively, joystick control data could be processed and sent to the wheelchair controller without modification. In this case the powered-wheelchair responded to joystick inputs as if it was an unmodified wheelchair system. Software systems were constructed using methods discussed in [12]–[14]. Systems had three main levels: supervisory, strategic and servo control. These were similar to the levels and sensor systems described or used in [15]–[17].

Algorithms applied the following rules: (1) User remained in overall control. (2) Systems only modified the trajectory of the powered-wheelchair when necessary. (3) Movements of the wheelchair were smooth and controlled.

## III. EXPERIMENTS

### A. Description

In this research, experiments were performed in using a dead-time simulator able to make arbitrary time lags [18][19]. Users performed driving experiments and the effect of motion lag was evaluated in the driving experiments to reveal the threshold value of the time lag,  $t_{\min\text{lag}}$ . The magnitude estimation method [20] was used. Furthermore, a phase lag was obtained for each perceived time lag when the input frequency was switched. Perceived phase lag was obtained from the time lag value. The finally evaluated value was the perceived phase lag related to each input frequency.

D.A. Sanders is a Leverhulme Trust Senior Research Fellow with the Royal Academy of Engineering based at University of Portsmouth, School of Engineering, Anglesea Building, Anglesea Road, Portsmouth, PO1 3DJ, UK (e-mail: david.sanders@port.ac.uk).

M. Langner is a Senior Engineer at Chailey Heritage, Anglesea, Portsmouth, PO1 3DJ, UK (e-mail: mlangner@chs.org.uk).

A.Gegov is a Reader at the University of Portsmouth, School of Computing, University House, Winston Churchill Avenue Portsmouth PO1 2UP, UK (e-mail: alexander.gegov@port.ac.uk).

### B. Dead-time simulator

The powered wheelchair system was modeled as a dead-time system [4] because the system gain was 1 at any frequency so that wheelchair motion response was confined to being affected by the time lag. So the motion lag indicated dead-time. A dead-time simulator created a virtual time lag and delayed the commands to the powered wheelchair. Random inputs were generated to test the operation of the simulator using:

$$G_S(s) = \exp(-\Lambda s) \quad (1)$$

### C. Magnitude estimation

The magnitude estimation method [4] was used to provide a quantitative evaluation of the effect of motion lag and perceived lag time was delivered by this evaluation. The magnitude estimation method revealed the effect due to various stimulations. The method was stable for quantitative evaluation of psychological reaction such as the effect of motion lag. The magnitude estimation method was one of the sensory assessments used in ergonomics [4][21]. S.S.Stevens [21] showed that a human psychological value  $\psi$  related with certain magnitude of physical stimulus  $I$  had the relationship:

$$\Psi = kI^n \quad (2)$$

Where:

- $\psi$ : psychological value
- $I$ : magnitude of physical stimulus
- $k$ : gain
- $n$ : power index

Equation (2) is Stevens' power law [4]. In the work described in this paper, the stimulation physical value is time lag. The psychological value  $\psi$  was evaluated from relating driving motion lag and the effect of the motion lag.

### D. Method

The method was similar to that used by Toyoda *et al* [4] with a tele-operated surgical robot used for coronary artery bypasses.

1. The simulator was set to operate at 0.1 Hz.
2. A metronome made a sound at 0.1 Hz rhythm and the sound and simulator were synchronized.
3. The dead-time was varied across 10 levels in a random order as shown below in table I.
4. The magnitude of time lag effect was recorded based on the magnitude estimation method.

TABLE I

Level	1	2	3	4	5
Dead time (mS)	1	2	3	4	5
Level	6	7	8	9	10
Dead time (mS)	7	10	14	19	25

Test subjects driving the wheelchair were eight men and women in their early 20's.

### E. Results

Perceived lag time for 0.1 Hz input frequency was transformed to a common logarithm and showed a linear curve. A linear curve indicated the response after an operator perceived a time lag. An approximate line was drawn to estimate the relationship. This result indicates that the perceived lag time is 80ms when input frequency is 0.1Hz. Equation (3) shows the relation between dead-time more than 80ms and magnitude of time lag affection.

$$\Psi = 25 * I^{0.35} \quad (3)$$

When dead-time was less than 80ms, a driver did not perceive any motion lag. The perceived lag time was obtained when input frequency varied from 0.1 Hz to 1Hz and Stevens' power law was calculated for those input frequencies. The relation between the dead-time, magnitude of time lag effect and the approximate power function after a driver perceived the motion lag were considered.

For example, phase lag was considered once a driver had perceived a motion lag. The phase difference of the dead-time system is shown by (4).

$$\angle G(j\omega) = -\omega\Lambda \quad (4)$$

So, the phase lag was determined by  $\omega$  which was the input angular frequency when operator perceived the motion lag and  $\Lambda$  which was the perceived lag time.

Perceived lag time, perceived lag phase and power index are shown in table II.

TABLE II

Input frequency (Hz)	Perceived lag time (ms)	Perceived lag phase (deg)	Power index
0.1	80	-6	0.35
0.2	70	-7	.4
0.5	50	-9	.41
1	25	-8	.35

## IV. DISCUSSION

Perceived lag time was obtained by the quantitative evaluation of the effect of motion lag using a magnitude estimation method. Table I shows that perceived lag time differs for each input frequency; if input frequency increases then perceived lag time decreases.

This result coincides with sensory prediction which is that during slow movements it is difficult for an operator to perceive any motion lag [4].

After a driver perceived a motion lag, that is when dead-time was greater than  $t_{\min\text{lag}}$ , the magnitude of the effect of the



time lag changed. The power index after the perception was an average of 0.36 which had no relation to input frequency. From this result, it was concluded that perceived lag time differs between each input frequency but after perception the magnitude of the time lag effect has no relation to input frequency.

Because the power index was less than 1, the effect of motion lag becomes insensitive as motion lag increased.

Considering perceived phase lag, when phase lag was less than -7 deg, a driver perceived the motion lag. As perceived lag time became larger perceived lag phase remained small. The difference for perceived phase lag is smaller than the perceived lag time. These results relate to the phase lag of the dead-time system (4). A driver had a threshold of phase lag in perception and then perceived lag time could be obtained in inverse proportion to the input frequency.

The threshold of motion lag perception was related to the system's phase lag. Therefore if a powered wheelchair driver does not perceive motion lag, it is important that the control system is designed so that the system's phase lag is less than the perceived lag phase. In this research, the perceived phase obtained was about -7 deg.

## V. CONCLUSION

The relation between motion lag and the magnitude of a time lag effect has been investigated and described and the perceived time was obtained. Perceived time lag determined perceived phase lag.

A perceived threshold value of the time lag,  $t_{\min\text{lag}}$ , the "perceived lag time" appears to exist and was estimated. After a driver began to perceive the motion lag, the magnitude of the time lag effect increases in relation to Stevens' power law. In this research, the Stevens' power index was an average of 0.36.

Perceived lag time was different for each input frequency and perceived time lag became smaller as input frequency increased. Regardless of input frequency, a driver tended to begin to perceive the motion lag if the phase lag became less than -7 deg.

Variation of perceived phase lag was smaller than that of time. When phase lag exceeded perceived phase lag then a driver began to perceive the motion lag in driving.

Reliability of perceived phase lag depended on  $t_{\min\text{lag}}$ . To improve the reliability of the results, more drivers need to be tested with the powered wheelchair system.

## REFERENCES

- [1] S. Chang, et al, "Time Delay Analysis in Teleoperation System," Proc. of the 1999 IEEE International Workshop on Robotics and Human Interaction, pp. 86-91, 1999.
- [2] K. Kawamura, T. Tajima, J. Okamoto, and M. G. Fujie, "Development of Real-time Simulation for robotic tele-surgery," Journal of Japan Society of Computer Aided Surgery, vol. 7, no. 1, pp. 7-14, June 2005.
- [3] D. Sanders "Analysis of the effects of time delays on the teleoperation of a mobile robot in various modes of operation". Industrial Robot: an International Journal, 36 (6). pp. 570-584. ISSN 0143-991X 10.1108/01439910910994641, 2009.
- [4] K. Toyoda, J. Okamoto and M.G. Fujie "The correlation between perception of motion lag and phase lag in tele-operation robot system". Proc. of 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, February 21 - 26, 2008.
- [5] D.A. Sanders, M. Langner M and G.E. Tewkesbury, "Improving wheelchair-driving using a sensor system to control wheelchair-veer and variable-switches as an alternative to digital-switches or joysticks". *IND ROBOT*, vol. 37, no. 2, pp. 157-167, 2010.
- [6] W. Gao and M. Hinders, "Mobile robot sonar backscatter algorithm for automatically distinguishing walls, fences, and hedges". *INT J ROBOT RES*, vol. 25, no. 2, pp. 135-145, 2006.
- [7] D.A. Sanders, I.J. Stott and M.J. Goodwin, "A software algorithm for the intelligent mixing of inputs to a tele-operated vehicle. *J SYST ARCHITECT*, vol. 43 no. 1-5, pp. 67-72. 1997.
- [8] D.A. Sanders, J. Graham-Jones and A. Gegov, "Improving ability of tele-operators to complete progressively more difficult mobile robot paths using simple expert systems and ultrasonic sensors". *IND ROBOT*, vol. 37, no. 5, pp. 431-440. 2010.
- [9] D.A. Sanders, Analysis of the effects of time delays on the teleoperation of a mobile robot in various modes of operation. *IND ROBOT*, vol. 36, no. 6, pp. 570-584. 2009.
- [10] D.A. Sanders *et al.* Simple expert systems to improve an ultrasonic sensor-system for a tele-operated mobile-robot. *SENSOR REV*, vol. 31, no. 3, pp. 246-260. 2011.
- [11] D.A. Sanders, Comparing ability to complete simple tele-operated rescue or maintenance mobile-robot tasks with and without a sensor system. *SENSOR REV*, vol. 30, no. 1, pp. 40-50. 2010.
- [12] G.E. Tewkesbury and [20] D.A. Sanders, "A new simulation based robot command library applied to three robots". *J ROBOTIC SYST*, vol. 16, no. 8, pp. 461-469. 1999.
- [13] D.A. Sanders, "Recognizing shipbuilding parts using artificial neural networks and Fourier descriptor, *PI MECH ENG B-J ENG*, vol. 223, no. 3, pp. 337-342, 2009.
- [14] G.E. Tewkesbury and D. Sanders, "A new robot command library which includes simulation". *J ROBOTIC SYST*, vol. 26, no. 1, pp 39-48. 1999.
- [15] D.A. Sanders, "Comparing speed to complete progressively more difficult mobile robot paths between human tele-operators and humans with sensor-systems to assist". *ASSEMBLY AUTOM*, vol. 29, no. 3, pp. 230-248. 2009.
- [16] J. Bergasa-Suso, D.A. Sanders and G.E. Tewkesbury, "Intelligent browser-based systems to assist Internet users". *IEEE T EDUC*, vol. 48, no. 4, pp. 580-585. 2005.
- [17] D.A. Sanders and A. Baldwin, "X-by-wire technology", *Total Vehicle Technology Conference* pp. 3-12. 2001.
- [18] K. Toyoda, et al, "Dexterous master-slave surgical powered robot for minimally invasive surgery-Intuitive interface and interchangeable surgical instruments", Proc. of 20<sup>th</sup> International Congress and Exhibition, Computer Assisted Radiology and Surgery CARS 2006, pp. 503-504, 2005.
- [19] K. Toyoda, et al, "Development of heart-beat compensation surgical robot for totally endoscopic Off-pump Coronary Artery Bypass Graft," Proc. of the 17<sup>th</sup> CISM-IFToMM symposium, pp343-350, 2008.
- [20] T. Fukuda, et al, Ergonomics guide, Scientist Inc. 2004.
- [21] S.S. Stevens, "On the psychophysical law," *Psychological Review*, 64(3), pp.153-181, 1957

# Building a Smart Homecare Supporting System for Elderly People Base on xBeacon Sensors

Lun-Ping Hung<sup>1</sup>, Tzu-An Pan<sup>2</sup>

<sup>1,2</sup>Department of Information Management, National Taipei University of Nursing and Health Sciences  
No.365, Ming-te Road, Peitou District, Taipei, Taiwan, R.O.C.  
e-mail: 640020163@ntunhs.edu.tw

**Abstract** - *In recent years, the average life expectancy of the global population has increased, meaning that the population structure is ageing. Hence, how to provide a good care environment for the elderly is a topical subject. In general, the elderly suffers from memory decline, thus often forgetting where they place their important items, and leading to life inconvenience. Moreover, more and more elderly live alone or live with their spouse, without their children at side. In case of emergency or illness, they are unable to find their drugs in time, which may result in life threatening situation. As the Information and Communication technology can effectively improve the homecare service for the elderly, this study proposes a hybrid signal strength cutting application and uses xBeacon to construct the home-based object sensing mechanism for the elderly. The proposed mechanism includes xBeacon configuration, received signal strength indicator, and visualized interface display. In the simulation of an elderly' home environment, the reader receives different strength values from the sensors at different positions, and the possible range of the object is cut according to the strength value of each mobile node by cutting node signal strength and zoning among the sensors, thereby deducing the position where the items are placed and the important items are tracked. The proposed mechanism can achieve the real-time sensing function and maintain independent and healthy life for the elderly.*

**Keywords:** Information and Communication, xBeacon, received signal strength indicator, elderly homecare

## 1 Introduction

In light of the global ageing trend, long-term care for elderly is a topical subject. As there is a growing demand for elderly long-term care, information technology has been used to develop novel devices that can share the workload of nursing staffs. For example, an accurate tracking model can be used to observe the daily behavioral model of the elderly, thus improving the care quality. Apple and Qualcomm have respectively launched indoor positioning service, iBeacon and Gimbal, which provide indoor mobile positioning services with smart device services built in simple structures. As the

technologies of the positioning devices become mature, they can be combined with household products to develop a new home application method [1].

Different countries have different criteria of assessing the health of the elderly. The report [2] on healthy elderly proposed by the World Health Organization (WHP) in 1989 suggested that activity of daily living (ADL) is an important indicator of the elderly life satisfaction. According to the statistics of Taiwan's Ministry of Health and Welfare, as of June 2013, the proportion of the population aged 65 and over who live alone and only live with spouse shows an increasing trend, with a population increase from 28% as of June 1990 to 31.7% in 2013. Many elderly need to take care of themselves. However, due to aging, their physical functions, ADL and memory gradually decline. They often forget where they place the important items (pill box, dentures and glasses box) in their houses. They may have to spend much time searching through the house. Inability to find important items may lead to inconvenience in life and emotional problems, or even physical illness in case of severe situations or emergencies.

eMarketer indicated that the number of smartphone users reached 1.4 billion in 2013. It is estimated that this number will exceed 1.76 billion by the end of 2014. Smartphones have become indispensable products of daily use. There are many apps for smartphones, among them, positioning service can be used to locate accurate position of users, and provide some smart service [3][4], such as Location Based Service (LBS) which was first used for military use. LBS uses satellite positioning technology in navigation and positioning, and integrates GIS, positioning technology and the network communication technologies. It can provide online location service on mobile devices. The mobile devices can obtain spatial location data or sensed values through built-in sensing device or location device (GNSS) for data calculation, analysis and decision making. The results are shown on the computers and mobile devices. In addition, some studies have used LBS for indoor location service [5][6]. This is the current development tendency, and can facilitate homecare tracking.

In indoor positioning, real-time information can be collected from various radio systems (Bluetooth, Wi-Fi) and

sensors (gyroscope, accelerometer) for more precise positioning. The portable devices combined with a variety of sensors have become the main development direction of indoor positioning. For example, Qualcomm developed the sensor Gimbal based on Bluetooth Smart technology. Apple developed Wi-FiSLAM, which applies Wi-Fi, GPS, gyroscope, accelerometer, and magnetometer for positioning. Apple also introduced the iBeacon sensing device with Bluetooth 4.0BLE, which provides indoor location service through Internet and GIS. It can send information to the smart phones. The Major League Baseball (MLB) also uses iBeacon technology to develop an app called "At the Ballpark" to receive iBeacon positioning information, providing a variety of services, including such as online shopping, baseball field video, e-tickets and position guide. Changed along different user locations, the content displayed on the app is automatically updated so as to provide real-time information for the users. However, after iBeacon Bluetooth technology is combined with smart devices, it is more suitable to provide indoor location-based service.

iBeacon is an indoor positioning system developed by Apple [7], which uses BLE technology [7]. BLE consists of Bluetooth v4.0 core and has low duty cycle, message length and rearrangement of division of work on the upper and lower layers, allowing the low energy bluetooth to have low energy consumption and low production costs. When the mobile device is located within the sensing range of the sensor, smart phones or the device with Bluetooth communication function can be used to perform the corresponding command. Each sensor has a fixed ID to identify each iBeacon. Compared with the traditional sensors, one button cell battery can allow the sensor to operate for 1-3 months. Due to low power consumption and convenient configuration, it has been widely applied. This study uses xBeacon and supports the apps developed by different developers. The tests showed that signals can be received and sent within 20 meters. Under normal operation, the button battery can supply power for more than one year. The sensor is as small as a coin, and is applied to indoor environment. As the operation time of xBeacon is long, it is used as a sensor in the object sensing system [8].

In order to effectively help the elderly to find their important items in their houses, this study developed a low energy bluetooth device, using node signal strength cutting and the sensors to find items such as pill box, glass boxes, dentures, etc. Based on the search records of the indoor home-based system and the item search situations, the elderly can observe the system functions which can help them to find the important items. The homecare environment for the elderly can be improved further to provide complete intelligent services with information technology, thus improving the quality of life for the elderly.

## 2 Methodology

The item sensing system for the elderly uses xBeacon to send signals. This study used the node signal strength cutting and sensor signal cutting method, and compared and cut between the nodes and sensors using RSSI. The positioning technology plays a very important role in this system, and is described in detail in this section.

### 2.1 The xBeacon Positioning Method and RSSI Positioning Method

With the advancement in technology, the indoor positioning system has been applied widely. Academically, the indoor positioning technologies have attracted continuous attention, and new technologies and algorithms have been improved and corrected. Real Time Location Systems (RTLS) as an immediate addressing mode uses received real-time information for positioning and tracking. In wireless network environment, immediate addressing mode can be divided into TOA, TDOA, AOA and RSSI [9]. The TDOA and AOA use signal time difference for positioning. This method needs accurate arrival time, and has great impact on the environment, making the general sensors unsuitable. The TOA uses signal arrival angle to determine the target location. The angle judgment requires the equipment of special specification. This may cause high equipment costs, and the signal may have error due to refraction. RSSI uses the received signal strength value as a basis for judgment because it does not need any special device. The configuration is easier than other positioning methods, and thus, is more widely applied to indoor positioning. In combination with various algorithms, it has been extended into many different positioning methods and can be classified into three types: Triangulation including Ring Overlapping Circle RSSI (ROCRSSI), Proximity including Max, and Scene analysis including Fingerprint and k-nearest neighbor (KNN) [10] [11] [12].

Shibo He proposed "wireless rechargeable sensor networks" [13]. Through the energy harvesting technology, the power of radio frequency, wind, solar energy and temperature collected by the sensor is converted into DC electricity via the converter, so as to make the sensor run [14] in the environment. Yuanchao Shu proposed a method to cut and partition nodes and estimate sensor location based on the different time of charge (TOC) because the sensor distance may cause different recharging time under WRSN framework [15]. This method has an advantage that the precise positioning can be achieved without arrangement of many sensors in the environment. The TOC positioning methods can effectively reduce number of sensors used in the environment but the rechargeable positioning sensors must be customized separately. In the environment, a wireless charger must be carried for the sensor. Upon related works, this study proposes a positioning method that is more suitable for the elderly based on the positioning method proposed by Shu. It

uses the node signal strength difference and signal node strength difference between the sensors in the environment and received signal strength value to compare the nodes, as well as shorten and cut the xBeacon location so as to locate the position of important items for the elderly.

### 2.2 Node signal strength cutting method

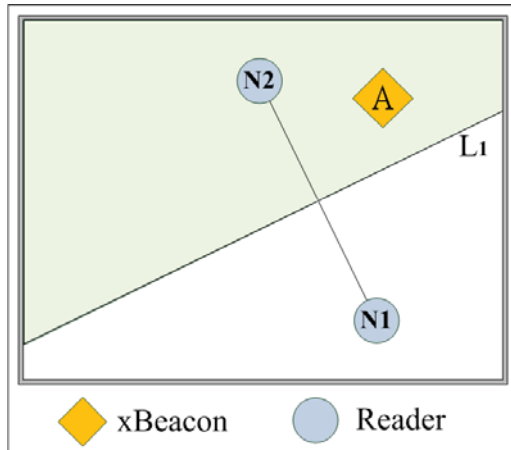


Fig.1. Regional cutting of two mobile nodes

Suppose sensor  $W$  is randomly distributed in the region  $R$ .  $R_n$  defined in this section denotes the region where each sensor is located. Each sensor is within its  $R_n$ . Each sensor is located in any place of  $R$  when there is no mobile node. This experiment uses vertical bisector  $L_n$  between the different mobile nodes to narrow range of the possible positions of the sensors, so as to obtain more accurate positioning results by narrowing the range.

In order to determine the location of the sensor, the mobile node is the position where the reader stops. When the reader stops in the environment region, the reader may collect information of the sensors within the range and average the received sensor strength values as the comparative data of each mobile mode. As the information of sensor  $A$  received by the mobile node  $N_1$  and  $N_2$  has different distance, the stored comparative data may also differ.

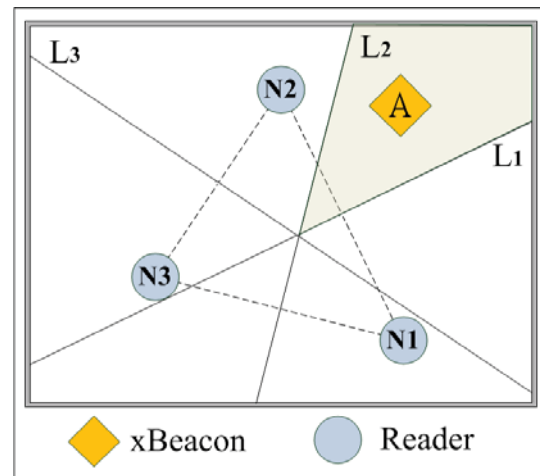


Fig.2 Regional cutting between three nodes

As shown in Fig. 1, sensor  $A$  is placed within the region, and is marked with yellow diamond with letter  $A$ . The reader randomly stops at two different positions marked with a numbered circle. The signal strength value of sensor  $A$  is received when the reader stops. The perpendicular bisector  $L_1$  between the two nodes can cut the area  $R$  into two parts  $R_a$  and  $R_b$ . In the mobile nodes  $N_1$  and  $N_2$ ,  $N_2$  strength value is greater than  $N_1$  because sensor  $A$  is closer to  $N_2$ . Thus, it can be known that sensor  $A$  location is within the left side  $R_a$  of perpendicular bisector  $L_1$  between  $N_1$  and  $N_2$ . As shown in Fig. 2, we suppose the reader moves, and stops at another position in the range. From the three nodes  $N_1$ ,  $N_2$  and  $N_3$ , it can be seen that the three perpendicular bisectors  $L_1, L_2, L_3$  cut the regional range, and can further narrow range of the possible position of sensor  $A$ .

Suppose  $r(A, N)$  is signal strength value of sensor  $A$  to  $N$ . When the sensor stops at the node  $N$ ,  $d(N_i, N_j)$  is the distance from  $N_i$  to  $N_j$ . It can be concluded that: if there is one sensor and mobile nodes  $N_1$  and  $N_2$  and  $r(A, N_1) \leq r(A, N_2)$ ,  $d(A, N_1) \leq d(A, N_2)$ . From this, by cutting the mobile nodes, possible range of all the sensors is convex polygon. As the mobile nodes  $N$  are different,  $\frac{N(N-1)}{2}$  perpendicular bisectors can be listed. For more and more mobile nodes, quantity of the perpendicular bisectors will be doubled to effectively narrow range of possible location of the sensors.

### 2.3 Reader signal strength cutting

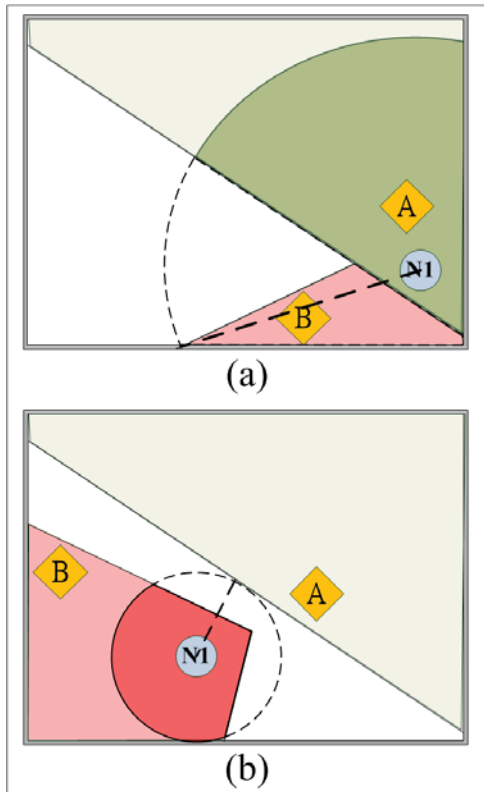


Fig.3 Schematic diagram of sensor signal strength cutting

This section describes how to narrow possible range by cutting reader signal strength.

As shown in Fig. 3 (a), suppose the sensors A and B are placed in the environment, and the two sensors may be located in  $R_a$  and  $R_b$ . After  $N_n$  receives A and B signal strength values, if  $N_n$  is located within  $R_a$  and A strength value is greater than B, it can be deduced that A may be located within the circle with center  $N_n$ , and the circle radius is the maximum distance from  $N_n$  to  $R_b$ . As shown in Fig. 3(b), if  $N_n$  is located within  $R_b$ , and A strength value is greater than B, it can be deduced that B may be located within the circle with center  $N_n$ , the circle radius is the minimum distance from  $N_n$  to  $R_b$ .

Suppose  $dis_{min}(.,.)$  denotes the minimum distance, as shown in Fig. 3(b). If the signal strength value of sensor B is smaller than sensor A, from the above, the sensor B is located outside of the circle area whose center is  $N_1$  and  $r = dis_{min}(N_1, R_b)$  because  $R_b$  shape may not remain convex polygon after cutting reader signal strength. To facilitate the following calculation and analysis,  $R_b$  is replaced with range of convex polygon through convex hulls (Fig. 4).

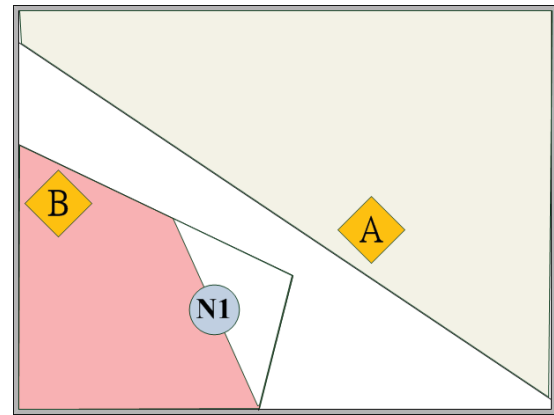


Fig4 Convex polygon range after processing

## 3 System structure and development

This section describes a home based sensor system for elderly which built in home indoor environment and uses the flow of finding the important items. It also introduces xBeacon positioning method to use signal strength value for comparison to narrow indoor range of xBeacon.

### 3.1 The system environment framework

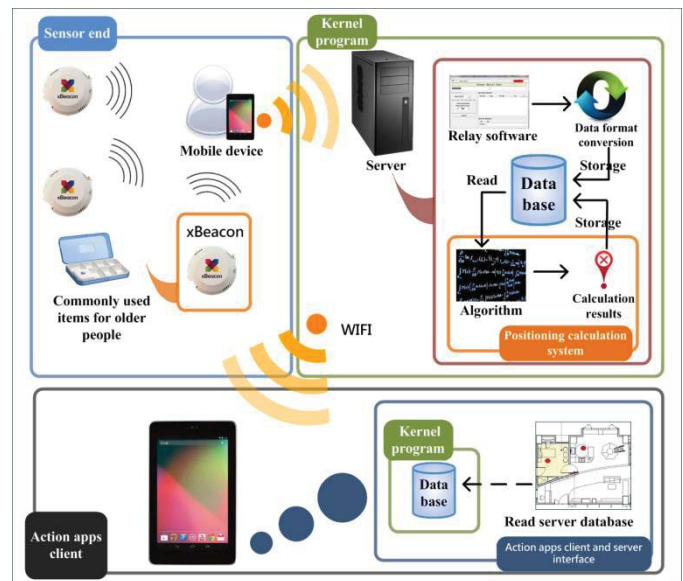


Fig.5 Flow chart of object position sensing system for the elderly

As shown in Fig 5, the home-based object position sensing system for elderly is divided into three parts: xBeacon sensor for sending information, core programs for information operation and processing and action apps for display image results. The mobile devices carried by elderly can receive signal information of xBeacon sensor within the range. The sensors in the environment can be divided into two types: one is installed in the fixed position in the environment

to assist in positioning location of elderly, and the other is installed on the important items to find the position of these items. The mobile devices receive signal strength of xBeacon and the corresponding number of the xBeacon in the areas, and these areas include a bedroom, kitchen, living room, bathroom and dining room and other indoor living environment of the elderly.

The core program can be divided into the relay software for receiving information, the database for storing information and positioning program for calculating information. By using sensor information sent from the mobile devices via WIFI, the relay program can classify data by different sensor purposes and convert the signal strength data into the format stored by the database. After the data processing, the data can be written into the database. The database stores ID number of each xBeacon, the signal strength and corresponding information of each important item. The relay software can write the received information into the database. The positioning calculation system can read the information in the database to identify positioning and classify and cut the possible locations of the sensors based on the strength values received by the sensors. After completion of positioning, the positioning calculation system may update the results to the database. The action apps receive the range of the sensors after calculation, visualize the range on the mobile phones, and mark position of the important items on the indoor map.

### 3.2 Development Tools

The server middleware of this system is developed using C++. The xBeacon positioning system uses Java EE platform for development. For the dynamic display of mobile phones and received data of xBeacon, the Eclipse JAVA is used as development tool and SVG is used for dynamic graphical display of real-time position of important items of the elderly and search confirm window. For the backend database, MySQL is used to access xBeacon signal records. The sensor information corresponding to the items, calculation results, item searching time record and relevant data for system, are provided for subsequent observation and analysis by administrators.

### 3.3 The Display of Information Monitoring System

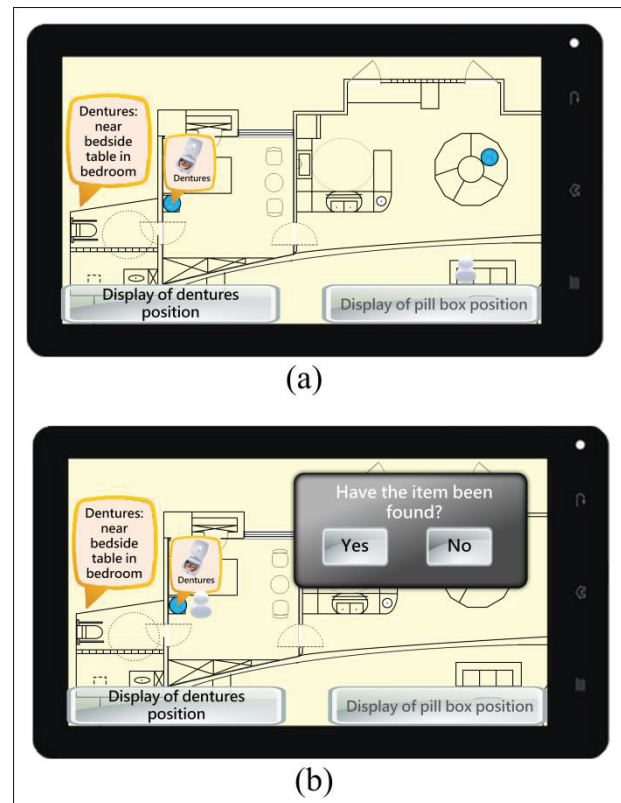


Fig.6 Small item searching system of mobile devices

In order to care for the elderly and search their lost items, this study uses portable devices to provide item search service for the elderly. If they want to search items, they only need to start the mobile apps and click the Item button. The application programs provide the item position according to the record of the item position in the database. The apps not only display the item position on the screen, but also relative position of the items on the left upper of the screen by referring to the items position to help the elderly to search the items, as shown in Fig. 6 (a). When the elderly move to proximity of items, the query “Have the items been found?” appears on the screen, and after click the above option, the position and time will be recorded in the database, as shown in Fig. 6 (b).

On the administrator interface, the user query time and position can be checked, and the time of searching the items can be calculated. Based on the long-term accumulated data, the activity characteristics of the elderly and their habits of placing the items can be analyzed. In addition, the administrator can select parameters and date range on the interface. The system marks the different items with dots of different colors, and the item position range is displayed with different circles connected by solid lines and dashed lines..

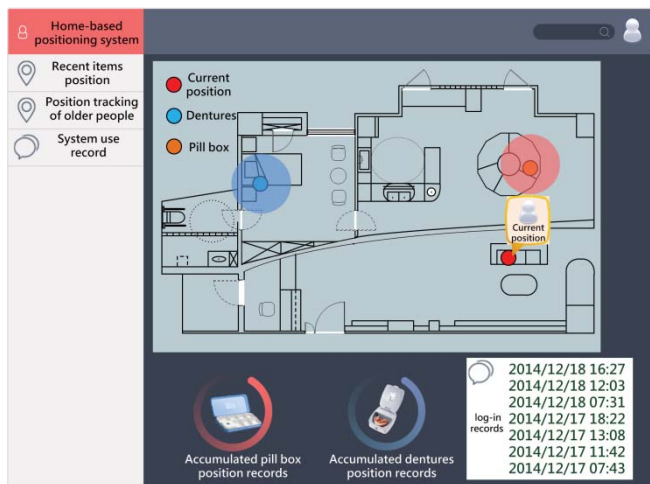


Fig.7 Administrator back-end interface

## 4 Conclusions

This study developed an object sensor system for the elderly to help them find the lost items in their indoor living environment. The proposed system uses the mobile devices to receive indoor xBeacon sensors in the indoor environment. In the indoor activities of the elderly, the mobile devices can automatically receive the information. After positioning the mobile devices, the mobile nodes are used to calculate range of xBeacon installed on the important items of the elderly, and record it in the database. The back-end administrator is provided to analyze action model and rules of the elderly. In the previous, RSSI was used with the fingerprint training database for positioning. Mobile node cutting is used to minimize quantity of the sensors deployed in the environment. The data of reference points are not required in the environment at first. The environment positioning is available when only fewer xBeacon sensors are deployed in the rooms, vitalized map is created. The system with simple deployment can be used in other environment. The smart devices has better portability and better data processing and transmission, and can effectively make the system integrate with sensors and information system in other home-based environment.

## 5 References

- [1] L. F. Purwoko, Y. Priyana, and T. Mardiono, "Ubiquitous Health Monitoring System design". 2013 Joint International Conference on Rural Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T), 1-6, 2013.
- [2] W. H. Organization, "Health of the elderly," 1989.
- [3] L. Kun-chan and S. Wen-Yuah, "Using Smart-Phones and Floor Plans for Indoor Location Tracking". IEEE Transactions on Human-Machine Systems, vol. 44, 211-221, 2014.

[4] K. Zhiqiang, T. Li, Y. Chongchong, and W. Zijun, "Research and implementation of intelligent mobile phone location based on RSSI in smart space". in Systems and Informatics (ICSAI), 2012 International Conference on, 1635-1639, 2012.

[5] B. Gressmann, H. Klimek, and V. Turau, "Towards ubiquitous indoor location based services and indoor navigation". in Positioning Navigation and Communication (WPNC), 2010 7th Workshop on, 107-112, 2010.

[6] C. Yuh-Ming, "Using ZigBee and Room-Based Location Technology to Constructing an Indoor Location-Based Service Platform". in Intelligent Information Hiding and Multimedia Signal Processing, 2009. IHH-MSP '09. Fifth International Conference on, 803-806, 2009.

[7] iBeacons Bible 1.0 [medium:online] Available: <http://www.gaia-matrix.com>

[8] M. Varsamou and T. Antonakopoulos, "A bluetooth smart analyzer in iBeacon networks". in Consumer Electronics Berlin (ICCE-Berlin), 2014 IEEE Fourth International Conference on, 288-292, 2014.

[9] Cisco Systems, Inc. (2008) Wi-Fi Location-Based Services 4.1 Design Guide - Location Tracking Approaches. [Online].<http://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/wifich2.html>

[10] Luo, Xiaowei, O'Brien, William J., & Ju-lien, Christine L. "Comparative evaluation of Received Signal-Strength Index (RSSI) based indoor localization techniques for construction jobsites". Advanced Engineering Informatics, 355-363, 2011.

[11] Ko, Chien-Ho. "RFID 3D location sensing algorithms". Automation in Construction, 19(5), 588-595, 2010.

[12] Altintas, B., & Serif, T. "Location-aware patient record access: Patient identification using fingerprinting technique". Proceedings of the 2011 IEEE International Symposium on a World of Wire-less, Mobile and Multimedia Networks, 1-6, 2011.

[13] S. He, J.-M. Chen, F. Jiang, D. Y. Yau, G. Xing, and Y.-X. X. Sun, " Energy Provisioning in Wireless Rechargeable Sensor Networks". Proc. IEEE IFONCOM, 2006-2014, 2011

[14] Ke L, Hao L, Chien-Chung S, editors. "Qi-ferry: Energy-constrained wireless charging in wireless sensor networks". Wireless Communications and Networking Conference (WCNC), 2012 IEEE, 2015-2520, 2012.

[15] S. Yuanchao, C. Peng, G. Yu, C. Jiming, and H. Tian, "TOC: Localizing wireless rechargeable sensors with time of charge". in INFOCOM, 2014 Proceedings IEEE, 388-396, 2014.





## **SESSION**

# **HEALTHCARE AND MEDICAL INFORMATION SYSTEMS, MONITORING SYSTEMS, DATABASES, BIG DATA, MEDICAL CODING**

**Chair(s)**

**TBA**



# Delivering Actionable Insights on Population Risk to Improve Health Outcomes and Reduce Healthcare Costs

J. Zhou<sup>1</sup>, M. Shepherd<sup>1</sup>, F. Li<sup>1</sup>, L. Fu<sup>1</sup>, J. Liu<sup>1</sup>, P. Liu<sup>1</sup>, L. Liao<sup>1</sup>, X. Wen<sup>1</sup>, J. Yao<sup>1</sup>  
D. Quebe<sup>2</sup>, J. Echols<sup>2</sup>

<sup>1</sup>Palo Alto Research Center, Palo Alto, CA, USA

<sup>2</sup>Xerox Corporation, Norwalk, Connecticut, USA

**Abstract** - *In this paper, we describe an approach to use data analytics, clinical domain knowledge, and automated workflows to deliver actionable insights to Medicaid programs and other overseers of population health. Modularized algorithms are developed in major areas to estimate population risk and evaluate the performance of Managed Care Organizations (MCOs). Clinical expertise orchestrates the algorithms into re-usable analytic flows that can answer specific and complex questions. The results of individual algorithms and/or the analytic flows can provide actionable insights and recommendations for improving health outcomes and reducing healthcare costs.*

**Keywords:** Machine Learning; Predictive Analytics; Big Data; Business Process Modeling; Workflow Orchestration

## 1 Introduction

There is a trend in U.S. State Medicaid offices to transition their members from a Fee-For-Service (FFS) payment model to a Managed Care payment model. This adds a level of separation between the providers and the states, relying on the Managed Care Organizations (MCOs) for inspection and improvement of provider services, introducing uncertainty on overall program performance. As a result, the Centers for Medicare and Medicaid Services (CMS) dictates that states provide better oversight of MCOs. Insights into patient data requires automated tools and processes for Medicaid directors and their staff to easily understand how each MCO is performing from clinical, financial, and operational perspectives. Prediction of future costs through population segmentation is insufficient for a Medicaid program to set policies and do budget management.

Beyond some risk assessment tools, most data analysis and workflow tools are manually driven. Clinical experts are required to make sense of the huge amount of data being generated through encounter claims. Staffs in Medicaid and other government healthcare programs will spend much of their time fulfilling reporting requirements and have difficulty attaining a better understanding of their member population. Data studies on population health often require domain knowledge and data science expertise. Data science provides the means to identify risk levels of specific populations that drive healthcare costs and reflect poor health outcomes.

Analytics are focused in core areas that are known to be major drivers of costs and outcome, such as Emergency Department (ED) Utilization, Hospital Readmissions, Chronic Disease Progression, and Access-to-Care. The analytics quickly identify specific populations that are at the most risk.

Clinical expertise provides the specific actionable recommendations to positively impact healthcare costs and health outcomes. The actionable recommendations are knowledge mined into automated intelligence for addressing the risks of populations identified by the analytics. The recommendations influence change in specific populations through various means, such as setting state policies, creating provider incentives, personalizing MCO assignment during member enrollment, offering outreach and education, and identifying opportunities for disease management initiatives.

## 2 Risk Assessment

The elderly and disabled, who constituted around 25 percent of the Medicaid population, accounted for about 70 percent of Medicaid spending on services in 2003 [1]. One way to reduce cost for state Medicaid is to identify high-cost patients and enroll them in more effective care management programs. There are existing risk models with some capability of predicting future cost based on claim data, including Chronic Illness and Disability Payment System (CDPS), Diagnostic Cost Groups (DCG), Adjusted Clinical Groups (ACG), Clinical Risk Groups (CRG), and Episode Risk Groups (ERG). Among them, CDPS is very popular in Medicaid program and is successfully used by states to adjust capitated payments for Medicaid beneficiaries.

The prediction accuracy of future cost was not the only goal when the model was developed; the requirements of payment systems were also taken into account such as stable payment amounts and resistance to proliferative coding. The predictive power of CDPS is not high for individual patient as shown by low prospective  $R^2$  (about 15% without truncation and 21% when truncated at \$100K) [2]. Other aforementioned models also suffer the same low  $R^2$  problem, for example,  $R^2$  is only 25% with truncation in DCG that has the highest accuracy. In general, these models are good at estimation of the aggregated healthcare cost of a population, but are not

suitable to identify potentially high-cost patients in a population.

Furthermore, identification of potentially high-cost patients has not always guaranteed the desired results. For example, a number of Medicare demonstration projects did not lower the cost even though the projects were able to identify high-risk patients [3]. The cost reduction and health outcome improvement can be achieved only if the effective interventions are implemented. The intervention probably is limited for a well-managed patient with multiple chronic conditions, while a frequent ED user with a single mild chronic condition can be managed better. The algorithms that stratify data by risk of future costs usually do not provide insight for actionable interventions. On the other hand, many potential outcome measures are strong indicators of care quality, or closely associated with high costs, or both. Therefore, rather than try to predict future cost, we concentrate on prediction of the outcomes whose improvement surely lead to cost savings and better care quality. We perform the risk stratification for patients by those important outcome measures instead of future cost.

### 3 Our Method

In Medicaid managed care programs, MCOs usually manage members, deliver healthcare services through contracts with third party providers, and pay for these services. To oversee the MCO operation, we examined member-centered outcome measures and provider-centered outcome measures as indicators of MCO performance and population health status. We have identified 4 major areas, including ED utilization, hospital readmission, chronic diseases progression, and access-to-care, which are the key areas where the clearest opportunities exist to reduce cost and improve outcome for Medicaid through the use of data analytics.

We took a hybrid data-driven and knowledge-based approach. Machine learning algorithms were used to build models in each area using 5 years of Medicaid claims in conjunction with clinical records and socioeconomic data. We also consulted with clinical experts to inject domain knowledge into model building process. These models can be used independently for answering simple questions or executed in sequence to form an analytic flow for answering complex questions in particular areas of concern. By leveraging domain knowledge, we can link these answers to actionable insights.

In the following sections, we will detail our approach to study each of the 4 major areas for improving health outcomes and controlling health care expenditures.

#### 3.1 Emergency Department Utilization

Between 2001 and 2008, the annual number of ED visits grew at roughly twice the rate of population growth [4].

Meanwhile, Medicaid beneficiaries use the ED at an almost two-fold higher rate than the privately insured [5]. With more beneficiaries gaining Medicaid coverage as a result of the Affordable Care Act, utilization of ED services is likely to increase further. Therefore improving appropriate use of ED services becomes our top priority. We have two goals: (1) Reducing the non-emergency ED usage (2) Reducing overall ED usage.

ED is supposed to be designated for truly emergent conditions. But inappropriate usage of ED contributes to about 20% of all the services, leading to overcrowding and waste [6]. In our analysis, we found the avoidable ED is related to availability of the alternative care. Urgent care centers, which are much cheaper than ED service, are recommended to mitigate the problem. In addition, children in 0-24 months are more likely to have avoidable ED than other age groups. And their most common conditions were upper respiratory infections and middle ear infections. This kind of ED encounters is avoidable with more pediatrician engagement.

Often an ED visit is the consequence of deteriorating health condition due to negligence in patient care. A skewed ED utilization pattern is found in Medicaid data we analyzed: only 8% of the total population used ED service; and for these ED services, over half of the cost was contributed by 20% of these members. Therefore, a more efficient approach to ED cost containment is to focus on populations with high ED risk. These patients accumulate large numbers of emergency department visits and hospital admissions which might have been prevented by relatively inexpensive early interventions and primary care. Predictive risk modeling is a relatively recent attempt to proactively identify this population [7] [8]. We take on the challenging task of building such models using claim data.

Our descriptive analysis has shown that the ED utilization is closely related to the patient health status, which is characterized by the historic ED visits, hospital admissions, and chronic conditions. Also ED utilization depends on the age and diagnosis. With the learned knowledge, we compiled three tiers of features for ED risk modeling. The first tier includes age group, gender, race, ethnicity, region, benefit program, and diagnosis code. These data are readily available from the claim dataset. The second tier are calculated from claims, including total medical cost, total pharmacy cost, number of ED visits, avoidable or not, number of hospital admissions and length of stay, chronic conditions, comorbidity, and access-to-care. The third tier contains other information that is not available in claim, such as socioeconomic status, life style, and mental status. Our current analysis only includes the first tier features and some second tier features as predictors.

The risk modeling is formulated by training with 1-year data then predicting the number of ED visits in the next 6 months. First we built a linear model for predictor selection. The backward stepwise selection reduced the feature size to 10. Various machine learning algorithms can be applied to

this regression problem. We have tested linear regression, neural network and random forest (RF). The RF gave the best prediction performance. Figure 1 shows the RF predictions vs observations. We can see about 80% of the variance is explained by this model. Meanwhile, the RF algorithm lists the importance of the predictors. It turns out that the previous ED count, ED cost, Age, and number of hospital admissions are among the most relevant predictors in our current model.

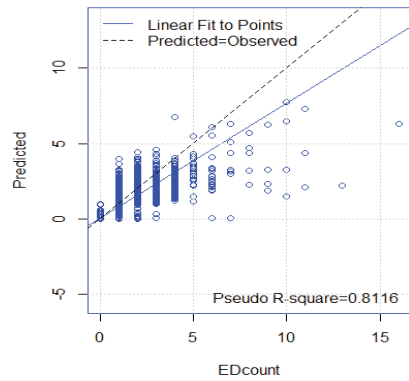


Figure 1. RF predictions vs observations

This preliminary risk model generates an ED risk score for each Medicaid beneficiary. With this score, states can risk adjust each patient and calculate the expected number of ED visit for each MCO. The difference between the observed and expected numbers is then used to gauge MCO performance. In the same way, they can compare FFS and MCO to tell if MCO is a better payment model. We are working on adding more tier 2 and 3 features and applying the model for patient cohort with certain condition, which will make our model more useful. For example, if we find out that access-to-care is the major driver for high ED utilization in one region, states can set the policy to force MCOs to increase provider coverage. Similarly if specific preventive care is a major driver, states can use incentives to encourage patients to increase compliance rate.

### 3.2 Hospital Readmission

Hospital readmission has been a key focus of quality improvement in health care systems. Many readmissions are linked with poor care quality during and after the primary hospital admissions, and are potentially preventable. These readmissions add huge burden on the patients and their families, and also are very costly financially. To improve health outcomes and reduce healthcare cost, CMS has been penalizing hospitals with high 30-day readmission rate in the Medicare program since 2012. In 2014, a total of 2610 hospitals received penalty, which accounts for about three quarters of all Medicare hospitals subjected to readmission reduction requirement.

There have been a lot of existing efforts on modeling hospital readmission [9] [10]. The focus has been mostly on predicting readmission risk for individual patients. The importance and value of individual readmission risk prediction is without doubt. The prediction can help hospitals and care managers to prioritize and improve the care

management for individuals. But perhaps more importantly, the hospitals, MCOs, as well as the states also need to understand the underlying drivers of readmission and potential actions to improve the outcome across the board.

We have developed a readmission prediction model with both goals in mind. In building the model, we work closely with clinical experts and emphasize on selecting and formulating features that are clinically meaningful and actionable. One example is that we took the Charlson Comorbidity Index (CCI) popular in existing models, and broke it down into a set of features on the individual Charlson conditions. It is difficult to define concrete actions on patients with high CCI, but much easier to develop and prioritize strategies to reduce readmissions if certain conditions such as diabetes with complications turns out to be a major contributor.

Table 1. Significant features in current readmission model

Primary inpatient	Length of stay
	Major diagnosis category
Demographic & administrative	Ethnicity
	Age
	MCO
Encounter history	Number of ED visits in past 6 months
	Number of admissions in the past year
	Number of readmissions in the past year
Comorbidities	Chronic obstructive pulmonary disease
	Liver disease
	Hemiplegia or paraplegia
	Malignancy and tumor

The readmission prediction model is formulated using logistic regression to predict the probability of all-cause 30-day readmissions. The definition of readmission here is that a patient being acutely re-admitted into hospitals within 30 days of discharge for any cause, except for pregnancy and psychiatric related inpatient stays. The model has been trained and validated with three years of claims on acute inpatient stays. Then we tested the model on claim data from two later years and obtained C-statistic between 0.77 and 0.8. This is comparable in discriminative power to the best models reported in the literature on similar data [9] [10]. The significant features in the final model are listed in Table 1. Certain interactions between the features have also been found to be significant.

From our current model, we identify a number of potential opportunities for reducing hospital readmissions for various population groups. The model indicates that one particular ethnicity is associated with higher readmission rate as shown in Figure 2. This is consistent with the results that our clinical expert found on a different state. This finding suggests directions of further investigation including network access, disease management, and member education for that particular population.

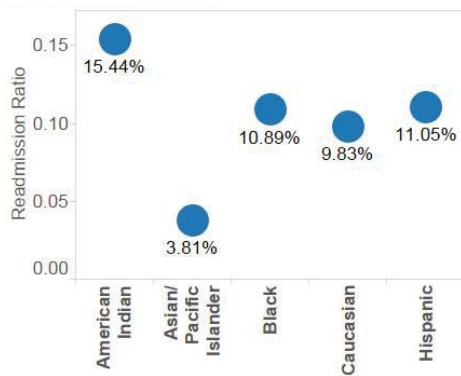


Figure 2. Readmission rates of different ethnicities.

Certain Charlson conditions appear to be significant in increasing readmission risk, including chronic obstructive pulmonary disease, liver disease, hemiplegia or paraplegia, malignancy and tumor. Potential actions might include medication review and specifically designed discharge instructions for members with these complications; increasing care coordination with primary doctors; member education on managing these conditions.

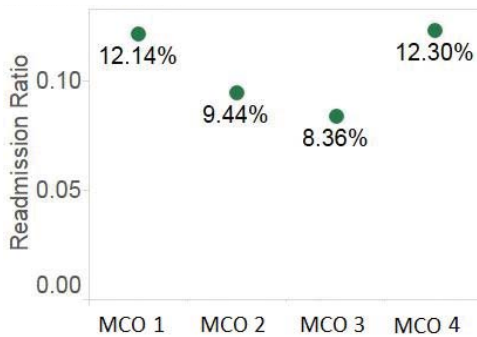


Figure 3 Readmission rates of different MCOs.

The MCO in which member enrolled also appear to be an important factor. In Figure 3, the difference of all-cause readmission rate between MCOs can be as much as 50%. The state can assign the members with a given condition to a particular MCO if this MCO does perform better than other MCOs in managing those members.

### 3.3 Chronic Disease Progression

Chronic disease is the leading cause of death and disability in the United States and accounts for 86% of the healthcare dollar spent [11]. Although it cannot be cured, chronic disease can be prevented and managed. However, to effectively manage the complexity of the chronic diseases, state Medicaid offices and MCOs need advanced analytics tools to help them predict future disease progression and costs, monitor the performance of the providers, and evaluate cost-effectiveness of care intervention.

Researchers have used the analytic models to predict the progression trajectory of diseases [12] [13]. The disease stages are mainly defined by clinical knowledge and the transitions between stages are modeled by Markov-based

models. Though progress has been made, there are still challenges: 1) Current models usually don't take into consideration of how long a patient has been in a certain disease state; 2) The initial state of a patient is hard to be identified with noisy observations; 3) It is complex to consider multiple covariates and model the interactions between different chronic diseases.

To mitigate the problem that health conditions might be under-reported in diagnosis codes [14], we determined the initial disease state of a patient in an observation window using observations beyond diagnosis codes, such as drugs, and treatment procedures, etc. Specifically, we used a random forest algorithm to classify diabetes into 3 states: pre-diabetes, diabetes without complication, and diabetes with complications. The classification accuracy on the test data is 94.5%. The top important features are the number of diabetes related outpatient visits, age, ED visits, and inpatient visits.

Next we used a semi-Markov model with consideration of time that a patient has stayed in a certain state. Because a patient's health condition could vary significantly depending on the patient's history, the conventional memoryless Markov model is too restricted for modeling disease progression. Instead we assume the sojourn time between two state transition events follows a Weibull distribution. The disease state transition is also influenced by other risk factors such as a comorbid disease, age, race, etc. The impact of these covariates is modeled using the Cox proportional hazard model:

$$\alpha(t|X) = \alpha(t)e^{\beta X} \quad (1)$$

where  $\alpha(t)$  is the baseline hazard rate defined by the Weibull distribution,  $X$  is the covariate vector, and  $\beta$  is the coefficient vector. Each patient's disease progression process is assumed to independently follow the semi-Markov process and the likelihood of the model is the log sum of likelihood of each individual process. Maximum likelihood estimation is applied to determine the optimal value for each parameter in the semi-Markov and Cox proportional hazard model.

We applied our model to modeling cardiovascular disease (CVD) progression with 3 disease states: pre-CVD, CVD, and mortality. A patient can transition from pre-CVD to CVD or mortality, and can transition from CVD to mortality. Three covariates are included in the proportional hazard model: age, gender and diabetes status, where diabetes status is a categorical variable indicating if a patient is in non-diabetes, diabetes without complication or diabetes with complication state at the time of the transition.

Separated models were trained for members in FFS program and MCO program. The hazard rates of developing CVD between FFS and MCO verses age for diabetes and non-diabetes members are shown in Figure 4. Hazard rate is the event rate at time  $t$  conditioned on the survival up to time  $t$ . As expected, FFS members have higher hazard rate than MCO members. Diabetes members also have higher hazard rate than non-diabetes members. The hazard rate also increases as a patient ages.

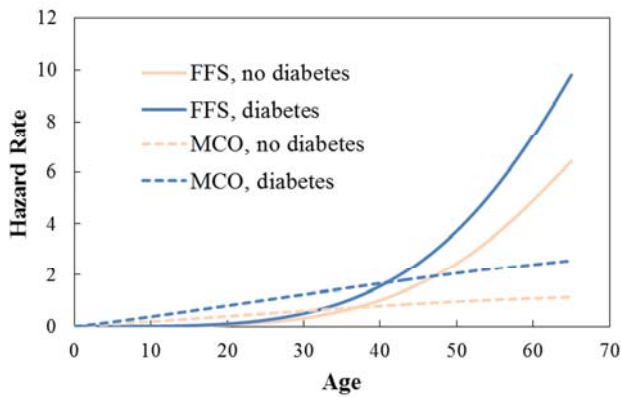


Figure 4. Hazard rates of developing CVD.

We use the model trained on claims from 2008 to 2011 to calculate the expected number of newly developed CVD members in the first quarter of 2012. Then we examined the real claims and counted the number of members who were actually diagnosed with CVD in the same period. The comparison between expected and observed numbers per 1000 members is shown in Figure 5.

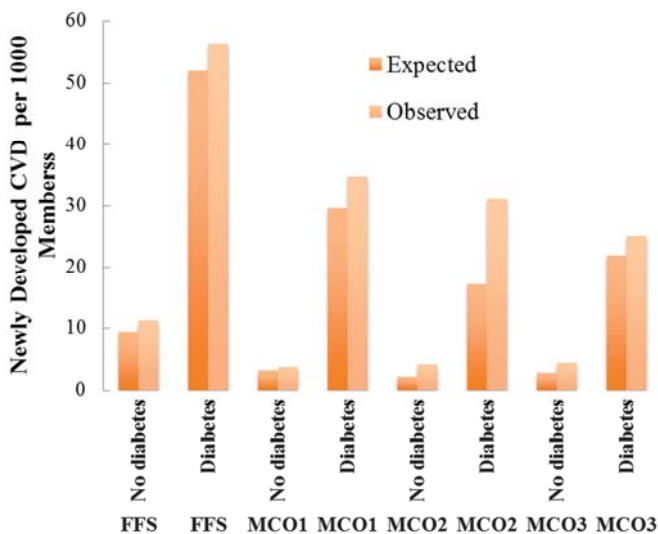


Figure 5. Prediction result of CVD disease progression.

FFS has significantly higher number than MCOs, which is expected because FFS members are usually in more severe health conditions. In addition, it shows that diabetes members have significantly higher risk to develop CVD than non-diabetes patients. To evaluate the performance of different MCOs, states can use the metrics that take into account the current member risk calculated from previous disease histories, e.g. the observed-to-expected new CVD member ratio. However there is a big deviation between the observed and expected numbers in one MCO, which is possibly due to the fact that other impacting covariates are not included in the model if not due to performance deterioration of this MCO. In addition to the 3 covariates we are currently using, we can add more features, such as comorbidities, utilization patterns, etc., into the covariate vector to improve our model accuracy.

The improved model can be used to identify the high-risk patients who are highly likely to develop a certain chronic disease in the near future. This will help MCOs to target these high-risk patients with early prevention and intervention. In addition, by projecting the future disease progression and costs without an intervention and comparing it with the observed outcomes after the intervention, the model can also be used to evaluate the cost-effectiveness of a disease management intervention.

### 3.4 Access to Care

Access-to-care is an important aspect of health care systems. Because it has a significant impact on the health outcomes and healthcare costs, states establish certain standards for access-to-care measure. Specifically for managed care programs in Medicaid, access-to-care is a measure of the capability of an MCO to provide its enrolled members with timely access to a sufficient number of in-network service providers and other health services that are covered in the benefit contract [15].

One important measure for access-to-care is network adequacy. To estimate network adequacy, the spatial distributions of services that the providers can supply, the demand that the members need, and the travel distance or time between members and providers must be calculated. For simplicity, we assume equal capacity among providers and equal demand from members. Figure 6 is one simple example of network adequacy analysis, where the color shows the percentage of members enrolled in a certain MCO whose access to endocrinologists failed to fulfill the state required geographic access standard. Several counties do not meet the state policy which is for 90% of the members to satisfy the standard.

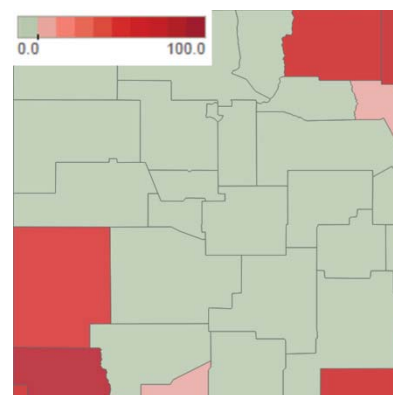


Figure 6. Counties with member percentage below standard.

The access-to-care model can measure access at very fine geographic resolution, and help states and MCOs identify regions with low network access. Furthermore, the model can recommend optimal locations for new network providers. Therefore, states can design incentives to encourage new providers towards these optimal locations. Also, MCOs can strategically add out-of-network providers into their networks to improve the access-to-care.

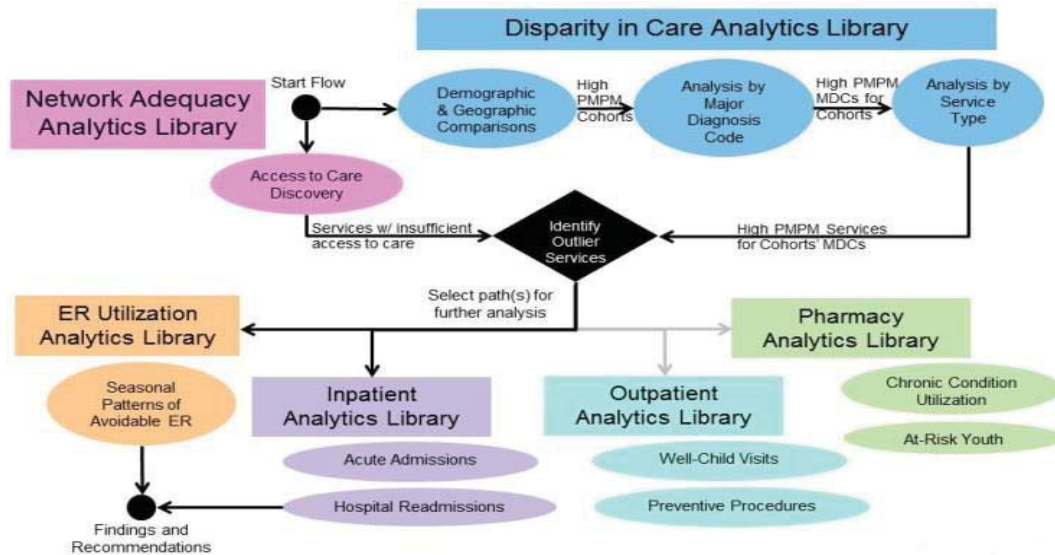


Figure 7. Examples of analytic flows

## 4 Analytic Flows

Through the use of the analytics for aforementioned major areas in combination with knowledge mined clinical expertise, an analytic flow framework [16] provides an approach to deliver actionable insights to state Medicaid offices and other overseers of population health. Through access to various data sources, the analytic algorithms are constructed in a modular fashion which can be connected together into analytic flows. In order to answer questions in particular areas of concerns, or drill down to find out root causes of high-level findings, a clinical expert can assemble the analytics into reusable flows [17] [18]. These flows may be re-executed on a regular basis as new data becomes available to the system, which provides a view into population health trends.

The findings in one analytic module may be passed to another analytic module for further processing. Also, new attribute values may be introduced which tailor the analytics in the flows to drive towards answering specific questions about the data.

An example analytic flow is shown in Figure 7 for assessing Medicaid per-member-per-month (PMPM) cost drivers through a demographic and geographic analysis of the population data. In this example, findings of diagnosis codes, access-to-care insufficiencies, and their service types lead to a decision point for further analysis. Dependent upon the findings at the decision point, additional analytics will be executed to determine the specific drivers of PMPM cost. In this example, over ED utilization is the driver. An analysis on different data using the same flow may show the lack of well-child visits is the driver.

## 5 Conclusion

We are developing analytic tools for state Medicaid offices to oversee the MCO operation to ensure high performance of Medicaid programs. Four major areas that have great potential of using data analytics to influence outcome improvements and cost savings were identified and discussed in this paper. Modularized algorithms to measure and/or predict outcome in each area were built and can be used to adjust the performance of MCOs in those areas for performance comparison. The different algorithms can be executed in sequence to form an analytic flow to answer specific and complex questions. The results of individual algorithms and/or the analytic flows with injected domain knowledge can provide actionable insights and recommendations. We have demonstrated the power of data analytics and the actionable insights it can provide to improve health outcomes and reduce healthcare cost.

We plan to add more features, such as socioeconomic status and lifestyle, into our models and improve our models with more realistic complexity.

## 6 References

- [1] The Kaiser Commission on Medicaid and the Uninsured, "Medicaid: A Primer – Key Information on the Nation's Health Coverage Program for Low-Income People," 2013.
- [2] F. Ross Winkelman, "A comparative analysis of claims-based tools for health risk assessment," April 20, 2007.
- [3] L. Nelson, "Lessons from Medicare's demonstration projects on disease management and care coordination," 2012.



- [4] A. B. Kharbanda, M. Hall, S. S. Shah, S. B. Freedman, R. D. Mistry, C. G. Macias, B. Bonsu, P. S. Dayan, E. A. Alessandrini and M. I. Neuman, "Variation in resource utilization across a national sample of pediatric emergency departments," *The Journal of Pediatrics*, vol. 163, no. 1, pp. 230-236, 2013.
- [5] T. C. Garcia, A. Bernstein and M. A. Bush, "Emergency department visitors and visits: who used the emergency room in 2007?" NCHS Data Brief, May 2010.
- [6] N. R. Hoot, L. J. LeBlanc, I. Jones, S. R. Levin, C. Zhou, C. S. Gadd and D. Aronsky, "Forecasting emergency department crowding: a prospective, real-time evaluation," *J Am Med Inform Assoc*, vol. 16, no. 3, pp. 338-345, 2009.
- [7] Z. Hu, B. Jin, A. Y. Shin, C. Zhu, Y. Zhao, S. Hao, L. Zheng, C. Fu, Q. Wen, J. Ji, Z. Li, Y. Wang, X. Zheng, D. Dai, D. S. Culver, S. T. Alfreds, T. Rogow, F. Stearns, K. G. Sylvester, E. Widen and X. B. Ling, "Real-time web-based assessment of total population risk of future emergency department utilization: statewide prospective active case finding study," *Interact J Med Res*, vol. 4, no. 1, p. e2, 2015.
- [8] W. Raghupathi and V. Raghupathi, "An overview of health analytics," *J Health Med Informat*, vol. 4, no. 3, p. 132, 2013.
- [9] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman and S. Kripalani, "Risk prediction models for hospital readmission: a systematic review," *JAMA*, vol. 306, no. 15, pp. 1688-1698, 2011.
- [10] C. v. Walraven, J. Wong and A. J. Forster, "LACE+ index: extension of a validated index to predict early death or urgent readmission after hospital discharge using administrative data," *Open Med*, vol. 6, no. 3, pp. e80-90, 2012.
- [11] National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP), "Chronic disease prevention and health promotion," Centers for Disease Control and Prevention, 2014. [Online]. Available: <http://www.cdc.gov/chronicdisease/overview/index.htm>.
- [12] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy and E. Couto, "Multistate Markov models for disease progression with classification error," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, no. 2, pp. 193-209, 2003.
- [13] M. F. Navdeep Tangri, M. M. F. Lesley A. Stevens, P. John Griffith, M. Hocine Tighiouart, M. Ognjenka Djurdjev, M. F. David Naimark, M. F. Adeera Levin and M. Andrew S. Levey, "A predictive model for progression of chronic kidney disease to kidney failure," *The Journal of the American Medical Association*, vol. 305, no. 15, pp. 1553-1559, 2011.
- [14] P. S. Romano and D. H. Mark, "Bias in the coding of hospital discharge data and its implications for quality assessment," *Medical Care*, vol. 32, no. 1, pp. 81-90, 1994.
- [15] J. Robert L. Phillips, E. L. Kinman, P. G. Schnitzer, E. J. Lindbloom and B. Ewigman, "Using geographic information systems to understand health care access," *Arch Fam Med*, vol. 9, no. 10, pp. 971-978, 2000.
- [16] S. M. S. d. Cruz, M. L. M. Campos and M. Mattoso, "Towards a taxonomy of provenance in scientific workflow management systems," in *Proc. of IEEE Congress on Services*, Los Angeles, 2009.
- [17] T. Oinn, M. Greenwood, M. Addis, M. N. Alpdemir, J. Ferris, K. Glover, C. Goble, A. Goderis, D. Hull, D. Marvin, P. Li, P. Lord, M. R. Pocock, M. Senger, R. Stevens, A. Wipat and C. Wroe, "Taverna: lessons in creating a workflow environment for the life sciences," *Concurrency and Computation: Practice and Experience*, vol. 18, no. 10, pp. 1067-1100, 2006.
- [18] W. Tan, J. Zhang and I. Foster, "Network analysis of scientific workflows: a gateway to reuse," *Computer*, vol. 43, no. 9, pp. 54-61, 2010.

# Evaluating a Potential Commercial Tool for Healthcare Application for People with Dementia

\*<sup>1a</sup> Tanvi Banerjee, <sup>2a</sup> Pramod Anantharam, <sup>3c</sup> William L. Romine, <sup>4b</sup> Larry Lawhorne, <sup>5a</sup> Amit Sheth

<sup>a</sup>Department of Computer Science and Engineering, Wright State University

<sup>b</sup>Boonshoft School of Medicine, Wright State University

<sup>c</sup>Department of Biological Sciences, Wright State University

<sup>1</sup>tanvi@knoesis.org, <sup>2</sup>pramod@knoesis.org, <sup>3</sup>william.romine@wright.edu, <sup>4</sup>larry.lawhorne@wright.edu, <sup>5</sup>amit@knoesis.org

Kno.e.sis Center, Department of Computer Science and Engineering, Wright State University  
303 Russ Engineering Building, 3640 Colonel Glenn Highway  
Dayton, Ohio 45435 USA

**Abstract** - *The widespread use of smartphones and sensors has made physiology, environment, and public health notifications amenable to continuous monitoring. Personalized digital health and patient empowerment can become a reality only if the complex multisensory and multimodal data is processed within the patient context, converting relevant medical knowledge into actionable information for better and timely decisions. We apply these principles in the healthcare domain of dementia. Specifically, in this study we validate one of our sensor platforms to ascertain whether it will be suitable for detecting physiological changes that may help us detect changes in people with dementia. This study shows our preliminary data collection results from six healthy participants using the commercially available Hexoskin vest. The results show strong promise to derive actionable information using a combination of physiological observations from passive sensors present in the vest. The derived actionable information can help doctors determine physiological changes associated with dementia, and alert patients and caregivers to seek timely clinical assistance to improve their quality of life.*

**Keywords:** Gerontechnology, activity monitoring, eldercare, patient monitoring, smart sensing

## 1 Introduction

Alzheimer's disease affects more than 5 million people claiming over 500,000 Americans annually [1]. As the sixth leading cause of death in Americans [1], its management is challenging. Current reactive healthcare costs more than 17% of GDP in the US [3, 4]. Alzheimer's related healthcare costs alone are around \$150 billion a year to Medicare and Medicaid [1]. To add to the challenge, dementia is an umbrella term that encompasses various forms of the disease such as Alzheimer's disease, vascular dementia, and

Huntington's disease, to name a few [2]. Not only are the healthcare costs associated with dementia staggering, but the impact on the caregivers is also a critical challenge; in 2013, 15.5 million family and friends provided 17.7 billion hours of unpaid care to those with Alzheimer's and other forms of dementia – care valued at \$220.2 billion [1]. With the exponential rise of the older population due to the baby boomers, the number of people with Alzheimer's disease (the most prevalent form of dementia) is estimated to reach around 13.8 million [1,6]. This creates the strong need for unobtrusive sensing modalities that can help monitor people with dementia and support caregivers.

With increasing adoption of mobile devices and low-cost sensors, an unprecedented amount of data is being collected [5]. However, in the context of dementia, it is challenging to convert this huge amount of data into actionable information that can: a) help detect behavioral changes in an individual with dementia and b) provide relevant information to the clinician supporting them in treating chronic illness. In our previous work, we derived actionable information from physical and physiological data collected from children diagnosed with asthma. We have developed kHealth kit [9, 28] a semantics-enabled smart mobile application with sensors, to capture observations from machine sensors (quantitative) and people (qualitative) in the domain of asthma [30]. We also have active clinical collaborations to investigate and evaluate the use of kHealth technology for reducing readmission of GI (gastrointestinal) and ADHF (acute decompensated heart failure) patients after their discharge from the hospital.

This paper reports our investigation to validate the sensors that will be used for the purpose for our ongoing study of physiological and behavioral markers of dementia. Using these markers, we aspire to build a model to detect changes in dementia patients and eventually attempt to predict adverse events. Specifically, we describe our preliminary work in

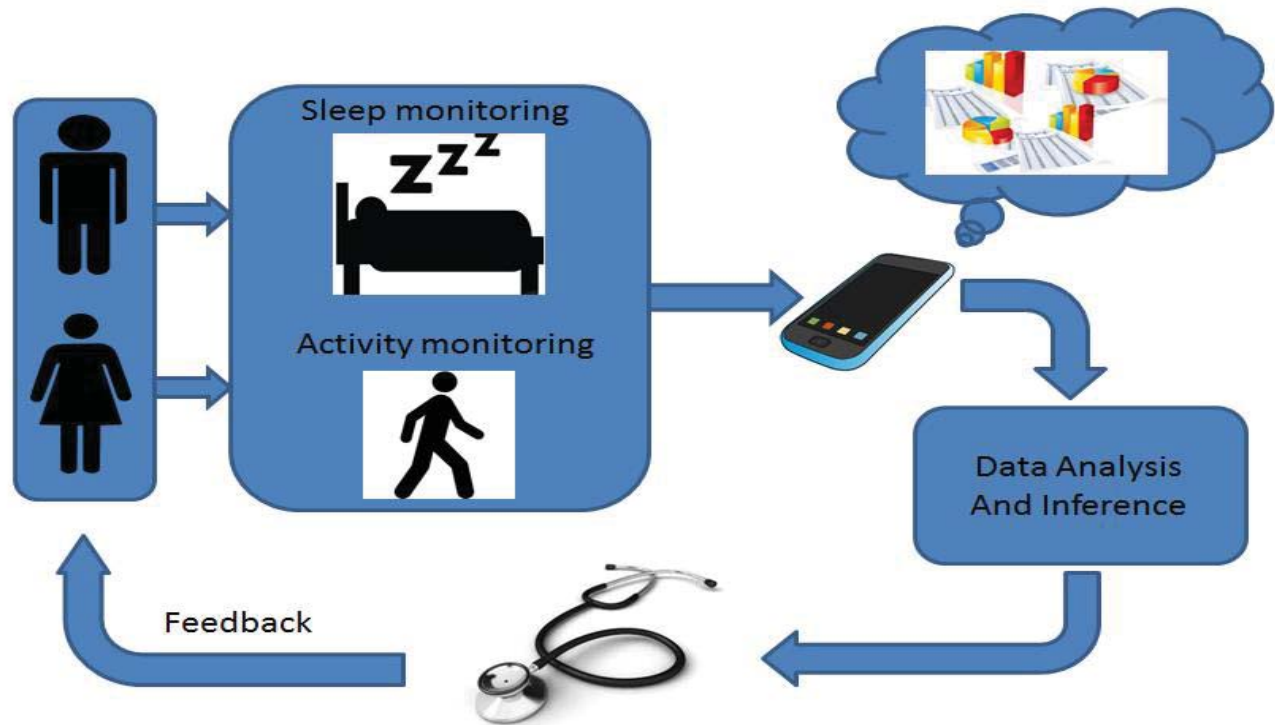


Figure 1. The kHealth Dementia application will measure physiological signals from the person with dementia as well as the caregiver to provide a deeper understanding of the behavior changes contributing to dementia

validating the parameters extracted from one of the most popular wearable sensors, Hexoskin, from Carre Technologies [7]. The Hexoskin system contains cardiac sensors, breathing sensors, and accelerometers that can be used to monitor movement, heart rate, and breathing in real-time [8]. Using this wearable technology, physiological parameters can be computed in a continuous and unobtrusive way. In this study, we analyze the data for six participants with varying demographics and discuss our results. Our analysis provides crucial insights into the physiological changes associated with dementia as well as analyze the temporal behavior of the patient. These insights can help clinicians diagnose and treat the illness and improve health management for the patients.

## 2 Related Work

The availability of low-cost sensors and mobile devices for monitoring physiological, physical, environmental, and cognitive health — within human bodies [10], on humans [11] and around humans [12] — are revolutionizing healthcare [13]. Microsoft Kinect and on-board sensors on mobile phones are being increasingly adopted in assisted living environments [14] and hospitals [15] for monitoring activities of daily living.

This trend is sure to accelerate with increasing FDA certification of devices or Internet of Things. With the rise of baby-boomers in the recent times, there has been an increase in research with Alzheimer's patients. In [16], the researchers conducted a study to compare the gait parameters for

dementia patients and older adults without dementia using a single waist-worn accelerometer. The results were promising; however, it may be difficult for a person with dementia to wear an accelerometer unless properly concealed. In other gait extraction methods, researchers tested gait parameters extracted using the Microsoft Kinect to validate the sensor data [17].

In recent work on Ambient Assisted Living (AAL) technologies, there has been a strong interest in commercially available sensors. Studies such as [18] have highlighted the potential for commercial sensors, including the Hexoskin vest as remote monitoring technologies that can help detect behavioral changes in older adults. However, there needs to be an assessment of the instrument before testing the sensor with dementia patients. In a recent study [23], the Hexoskin was used as a sensor to test the quality index of the ECG signal from the heartrate sensor. However, there was no evaluation of the parameters extracted from the sensor or instrument validation for the data obtained from the Hexoskin. There is also an unpublished work using the heartrate, breathing rate and activity sensors from the Hexoskin at [27]. However, there was no evaluation of the cadence parameter, nor exploration of the parameter relationships in the study that is essential before using this sensing modality for longitudinal studies in healthcare domains like dementia.

In this work, we present a systematic study and a much deeper understanding of various physiological observations collected from Hexoskin. We demonstrate the feasibility of

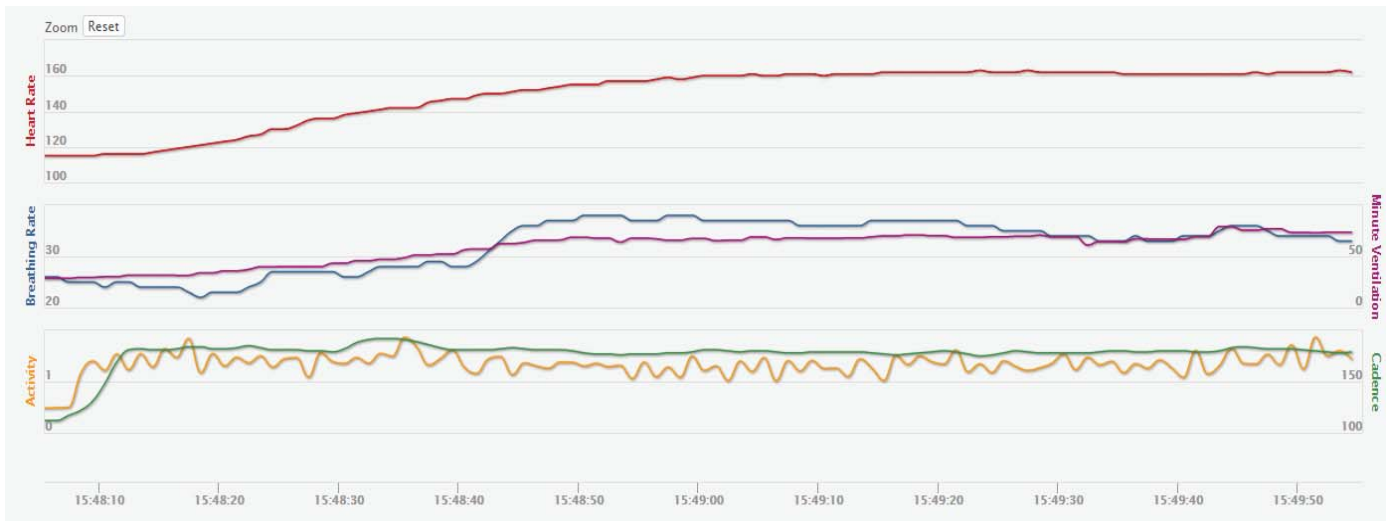


Figure 2. The Hexoskin data for a Run sequence showing the HR (top red), BR (blue middle), and A (yellow bottom) signals are labeled with the Y Axis on the left, and MV (purple middle), and C (green bottom) signals labeled with the Y Axis on the right. The X Axis represents time in the format hour : minute : second.

using insights gained by the analysis of physiological observations in dementia management.

### 3 Experiments and Analysis

In this section, we will discuss our proposed system setup for monitoring people with dementia. In the proposed plan of study, we hope to study the behavior of both the person with dementia as well as the caregiver. In the neuropsychiatric inventory study by Cummings [19], greater cognitive impairments were reported for people with dementia over time. The symptoms include agitation, apathy, depression, aberrant motor behavior, and abnormal nighttime behavior. Moreover, the deterioration in patient behavior and personality increases the stress on caregivers and leads to negative outcomes which need to be monitored to ensure that the caregiver can sustain his or her role to provide support to the person with dementia [20]. By continuous, unobtrusive monitoring of the physiological parameters of the person with dementia as well as the caregiver, we can detect changes in the movement and sleep patterns of the patient as well as the stress generated on the caregiver that are useful indicators for clinicians. Figure 1 shows the overall block diagram of our proposed approach. Commercially available sensors will be used to monitor both patients and caregivers, a daily questionnaire based on the Zarit Burden Interview questions [22] will be routinely asked to the caregiver using a mobile application to provide ground truth. We utilize statistical and machine learning approaches to extract behavioral patterns of the patients so that we can detect anomalies that can be used to predict behavioral disturbances in people with dementia. This information can then be provided to the clinician for further action on their part.

As mentioned earlier, one of the most promising sensors to monitor the patient is the Hexoskin vest. This measures five parameters: heart rate (HR) in beats per minute (BPM), breathing rate (BR) in BPM, minute ventilation (MV) to detect the volume of gas inhaled or exhaled by the lungs in lungs per minute (LPM), cadence (C), as well as the activity level (A) on a scale of 0 to 1 using accelerometers in the X, Y, and Z directions (resolution of 0.004g) [7]. We can see a sample of the data extracted using the Hexoskin in Figure 2. Figure 2 shows an example of a Run activity. We can see that the HR increases gradually over time during the participant's activity as we would expect. Similarly, we see an expected increase in the other parameters over time as the person's C, A, BR, and MV rises over time. This corroborates our understanding of the Run activity. We will look at a more in-depth statistical analysis of the parameters in Sections 3.1 and 3.2.

As compared to other commercially available sensors, this has the added benefit of being worn as an under shirt by the person with dementia, instead of wearable bracelets like the Fitbit [21] that could confuse the patient who could then possibly take it off. Moreover, the Hexoskin vest is Bluetooth enabled with over 14 hours of battery life and can locally store more than 150 hours of recording [7]. We will now discuss our instrument validation for the cadence measurement of the vest in Section 3.1.

#### 3.1 Cadence Validation in a Controlled Setting

In this subsection, we first validate the C (cadence) parameter since there is a direct relationship between the gait related activities mentioned above with cadence. Moreover, studies such as [16] highlight the importance of gait-based features in differentiating between people with dementia and people without. For this validation, there were four

participants, two male and two female, between the ages 30-35. Data were captured in a controlled method with each participant asked to sit for ten minutes, walk for ten minutes, run at their normal pace for ten minutes, and run hard (sprint) for one minute. Figure 3 shows the box plots for the C parameter for the four participants for the four activity states. As we can see from Figure 3, the fitness across the four participants varies; especially for the Run and Sprint activity states. As can be expected, the cadence value is zero for all four participants at rest. Moreover, we can see the variance for the different individuals varies for the different activity states. For example, Subject 2 has much lower variance for the Sprint activity state whereas Subject 4 has a high variance for the same activity state. We also see that there are several outlier values for the Run and Sprint activity states for Subject 1. This was due to the participant walking for very short intervals during the data recording.

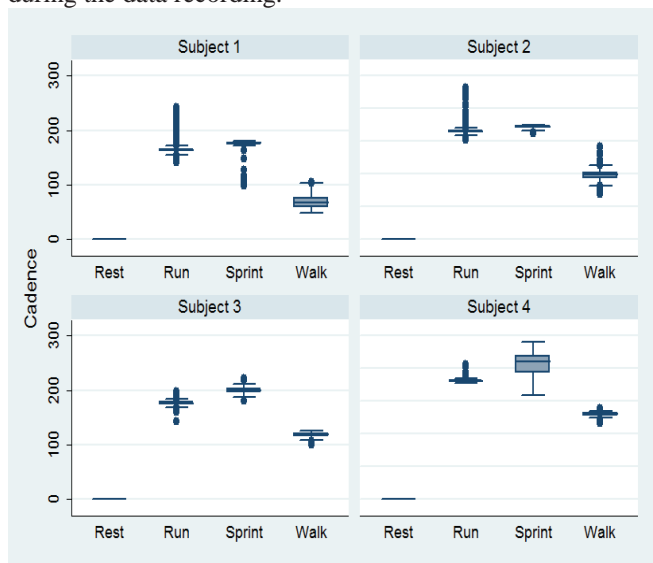


Figure 3. Box plots for Cadence (C) for the four participants in the controlled setting for the different activity states.

Table I shows the average C values and the standard deviation across the four participants for the Rest, Walk, Run, and Sprint activity states. Table I and Figure 3 collectively show the expected result that:

$$C_{\text{Rest}} < C_{\text{Walk}} < C_{\text{Run}} < C_{\text{Sprint}}$$

Table I. Cadence (C) Mean and Standard Deviation Results across Subjects for the Different Activity States

Activity State	Mean	Std. Dev
Rest	0.00	0.00
Walk	103.05	25.03
Run	171.95	10.25
Sprint	185.93	22.00

Using minimum norm quadratic estimation (MINQUE) [29], we explored the intra-class correlations of subject, activity state, and their interaction, with the data. We found that a majority (95.5%) of the variation in the parameter C can be explained by activity state while only 1.8% of the variation can be explained directly by differences between the subjects. 2.0% of the variation in cadence can be explained by individual differences in cadence between activity states. This leaves only 0.7% of the variation in cadence unexplained by Subject and Activity State. This indicates that the activity state affects the variance of C much more than differences in subjects or random error.

In summary, we find that differences in cadence between activity states are aligned with expectations. Further, we find that a vast majority of the variation in cadence can be explained by differences between subjects and activity states. These findings collectively support the precision and utility of the Hexoskin's C parameter for detecting changes in activities across individuals.

### 3.2 Physiological Parameter Evaluation in Semi-Controlled Setting

In this study, we evaluate the performance of all the parameters extracted from the Hexoskin: HR, BR, MV, C, as well as A. Six participants were asked to perform walks in semi-structured settings in accordance with their comfort level. The participants ranged from ages 30 – 65 and were all healthy adults. Two were female and the remaining four were male participants. As mentioned earlier, the only requirement was that the participants perform some gait-related activities according to their comfort level. This involved walking around the house, inside the house, performing household chores, as well as sitting and resting. Data were recorded for approximately 26 hours for this experiment. Since we validated the C parameter in Section 3.1, we will use this as the key measure to evaluate the performance of the remaining physiological parameters.

#### 3.2.1 Multivariate Analysis

In this subsection, we describe our results on the multivariate analysis of variance (MANOVA) with cadence (C) as the independent variable (IV) and the remaining parameters (A, MV, BR, HR) as the dependent variables (DV). In this way, we can see the effect of C on the remaining parameters as a whole instead of separately, as we see the individual relationships between the parameters in Section 3.2.2 [24]. We use the F-statistic calculated from Wilk's Lambda as a multivariate criterion [25] for statistical significance of C as a predictor of the other parameters together ( $\alpha = 0.05$ ) for each subject. Partial  $\eta^2$  is used as a measure of the percentage of variance in the DV's explained by C.

Table II. MANOVA results with C as IV and HR, BR, A, and MV as DV.

MANOVA	Lambda	F*	Partial $\eta^2$
Subject 1	0.128	28922.56	0.871
Subject 2	0.160	26888.12	0.839
Subject 3	0.181	32369.65	0.818
Subject 4	0.255	3275.61	0.744
Subject 5	0.375	8020.30	0.624
Subject 6	0.242	6354.81	0.757

\*Significant at  $\alpha \ll 0.001$ , p-value  $\sim 0$

As we can see in Table II, cadence is a highly significant predictor of the DV's in all subjects (since the p-value is  $\sim 0$  for all the six participants). Moreover, C explains between 62% - 87% of the variance for all the DVs across the six participants (as can be seen from the Partial  $\eta^2$  values). Since cadence is a direct measure of activity state, it makes sense that parameters associated with heart rate, respiration, and activity will be highly related to cadence. Specifically, it is an expected finding that as C increases, the subject's oxygen demand is going to increase, which will in turn result in an increase in breathing and heart rates. Positive correlations can be used as evidence that the Hexoskin is working as expected, and is taking valid measurements. Now that we have looked at the variance analysis for these physiological parameters as a group, we explore the individual correlations in the following section.

### 3.2.2 Correlations between parameters

In this subsection, we will look at the relationship between the five parameters from the Hexoskin sensor. Table III shows the average Pearson correlation [26] between the five parameters across the six participants. The parameters significant at 95% confidence interval are highlighted in bold.

Table III. Average Pearson Correlation Results between the Five Parameters across the Six Participants.

	Mean	Std. Dev	SE	$T_{df=5}$	P-value
<b>C-BR</b>	<b>0.54</b>	<b>0.20</b>	<b>0.08</b>	<b>6.53</b>	<b>0.001*</b>
C-HR	0.16	0.28	0.12	1.38	0.226
<b>C-MV</b>	<b>0.66</b>	<b>0.15</b>	<b>0.06</b>	<b>10.9</b>	<b>0.000*</b>
<b>C-A</b>	<b>0.85</b>	<b>0.07</b>	<b>0.03</b>	<b>28.9</b>	<b>0.000*</b>
BR-HR	0.18	0.28	0.11	1.56	0.180
BR-MV	0.18	0.21	0.09	2.04	0.097
<b>BR-A</b>	<b>0.52</b>	<b>0.18</b>	<b>0.07</b>	<b>7.06</b>	<b>0.001*</b>
<b>MV-HR</b>	<b>0.31</b>	<b>0.28</b>	<b>0.11</b>	<b>2.75</b>	<b>0.040*</b>
<b>MV-A</b>	<b>0.64</b>	<b>0.18</b>	<b>0.07</b>	<b>8.93</b>	<b>0.000*</b>
HR-A	0.19	0.28	0.11	1.69	0.152

\*Significant at  $\alpha = 0.05$

As we can see, six of the parameter pairs show a strong correlation at the 0.05 level. The parameter HR shows the least correlation among the parameters. It is the least significant predictor for cadence across the six participants. Many factors could lead to this: the different body shapes and sizes may lead to different placements of the heart sensor across the participants that could lead to an error in measurement. HR varies across subjects based on the fitness level and HR can change due to stress or caffeine intake (observed with one of our participants), which has no relationship with C. Also, since a person's heart is always beating (at varying levels) even when at rest, the HR value does not change as much as the other parameters when the person's C or A increases; leading to lower correlation. This could specifically be the case for the Walk activity since the HR parameter may not increase significantly. However, we find that the BR and MV parameters exhibit more variation with C; these show strong potential to complement C and A in the detection of behavior patterns.

## 4 Acknowledgement

We thank Wright State University's Vice President for Research for the partial funding of this effort. We also thank Vaikunth Sridharan for supporting kit preparation and all the participants of our study for collecting real-world data for our analysis.

## 5 Conclusions

We tested and validated the sensor data extracted from the Hexoskin vest. We first validated cadence using different activity states of rest, walk, run, and sprint in a controlled setting. We then evaluated the performance of the other parameters in a semi-controlled environment using six participants from a more diverse age group. The parameters BR, MV, and A were found to be consistent with the C values. These show strong potential to differentiate between different activity and behavior patterns. This shows that all parameters except for HR may be directly useful in detecting changes in patient behavior for future studies. Even the HR sensor may be a useful parameter if we look at other features like temporal differences spikes in HR. We plan to explore these temporal trends in our future experiments.

We build on kHealth's foundation to test our hypothesis that an evidence-based approach can help doctors determine more precisely the changes in behavior patterns for people with dementia. All our kHealth applications involve active clinical collaborations with medical professionals leading to evaluation with patients. Our next step is to test the system using participants with dementia. The Hexoskin has shown strong promise as a sensor platform for detecting changes in activity and behavioral patterns. Additional research is needed to study the efficacy of these physiological parameters as predictors for behavioral change in people with dementia. We can then develop a precise understanding of these effects in

dementia patients in order to quantify the sensed data's role for clinical assessment of their symptoms. The derived understanding can be used to alert caregivers and physicians so that appropriate measures can be taken to ensure the safety and well being of both the people with dementia, as well as the caregivers.

## References

- [1] Alzheimer's Association description of Alzheimer's statistics, Available online at: [http://www.alz.org/alzheimers\\_disease\\_facts\\_and\\_figures.asp#quickFacts](http://www.alz.org/alzheimers_disease_facts_and_figures.asp#quickFacts)
- [2] Dementia related facts, Available online at: <http://www.cdc.gov/mentalhealth/basics/mental-illness/dementia.htm>
- [3] D. A. Squires, "The U.S. Health System in Perspective: A Comparison of Twelve Industrialized Nations," June 2011, Available online at: <http://bit.ly/oZwhFZ>
- [4] Health Costs: How the U.S. Compares With Other Countries, Available online at: <http://www.pbs.org/newshour/runtdown/2012/10/health-costs-how-the-us-compares-with-other-countries.html>
- [5] Quantified Self <http://quantifiedself.com/>
- [6] G. K. Vincent, V. A. Velkof, "The next four decades: The older population in the United States: 2010 to 2050." Washington, D.C.: U.S. Census Bureau; 2010.
- [7] Hexoskin. [www.hexoskin.com](http://www.hexoskin.com)
- [8] A. Pasolini, "Sensor-packed Hexoskin shirt measures performance in real time". Available at: <http://www.gizmag.com/hexoskin-sensor-t-shirt-body-metrics/29098/> Gizmag, September 19, 2013.
- [9] kHealth: A knowledge-enabled semantic platform to enhance decision making and improve health, fitness, and well-being, Available online at: <http://knoesis.org/projects/khealth> (Accessed May 27, 2013).
- [10] V. Santhisagar, T. Ioannis, B. Diane, J. C. Faquir, P. Fotios, "Emerging synergy between nanotechnology and implantable biosensors: A review." *Biosensors and Bioelectronics* 25.7: 1553-1565, 2010.
- [11] J. S. Karlsson, U. Wiklund, L. Berglin, N. Östlund, M. Karlsson, T. Bäcklund, & L. Sandsjö, "Wireless monitoring of heart rate and electromyographic signals using a smart T-shirt." In *Proceedings of International Workshop on Wearable Micro and Nanosystems for Personalised Health*, 2008.
- [12] M. Chan, E. Campo, D. Estève, & J. Y. Fourniols, "Smart homes—current features and future perspectives." *Maturitas*, 64(2), 90-97, 2009.
- [13] E. Topol, "The creative destruction of medicine: How the digital revolution will create better health care." *Basic Books (AZ)*, 2012.
- [14] T. Banerjee, M. Skubic, J. M. Keller & C. C. Abbott, "Sit-To-Stand Measurement For In Home Monitoring Using Voxel Analysis," *IEEE Journal of Biomedical and Health Informatics*, 18(4):1502-1509, 2014.
- [15] T. Banerjee, M. Rantz, M. Li, M. Popescu, E. Stone & M. Skubic, "Monitoring Hospital Rooms for Safety Using Depth Images," *Proceedings, AAAI Fall Symposium Series - AI for Gerontechnology*, Washington DC, November 2-4, 2012.
- [16] M. Gietzelt, K. H. Wolf, M. Kohlmann, M. Marschollek, and R. Haux. "Measurement of Accelerometry-based Gait Parameters in People with and without Dementia in the Field." *Methods Inf. Med* 52, no. 4: 319-325, 2013.
- [17] E. Stone & M. Skubic, "Evaluation of an Inexpensive Depth Camera for In-Home Gait Assessment," *Journal of Ambient Intelligence and Smart Environments*, 3(4):349-361, 2011.
- [18] A. L. Bleda, R. Maestre, A. J. Jara, & A. G. Skarmeta, "Ambient Assisted Living Tools for a Sustainable Aging Society." In *Resource Management in Mobile Computing Environments* pp. 193-220. Springer International Publishing, 2014.
- [19] J. L. Cummings, "The Neuropsychiatric Inventory: Assessing psychopathology in dementia patients." *Neurology* 48 (Supple 6): S10-S16, 1997.
- [20] W. E. Haley, E. G. Levine, S. L. Brown, and A. A. Bartolucci. "Stress, appraisal, coping, and social support as predictors of adaptational outcome among dementia caregivers." *Psychology and aging* 2, no. 4: 323, 1987.
- [21] Fitbit sensor. [www.fitbit.com](http://www.fitbit.com)
- [22] S. H. Zarit, K. E. Reever, J. Bach-Peterson, "Relatives of the impaired elderly: correlates of feelings of burden." *Gerontologist*. 20:649-55, 1980.
- [23] D. Tobon Vallejo, T. Falk, M. Maier, "MS-QI: A Modulation Spectrum-Based ECG Quality Index for Telehealth Applications," *Biomedical Engineering, IEEE Transactions on*, 2015.

- [24] R. T. Warne, "A primer on multivariate analysis of variance (MANOVA) for behavioral scientists". *Practical Assessment, Research & Evaluation* 19 (17): 1–10, 2014.
- [25] K. V. Mardia, J. T. Kent, J. M. Bibby, "Multivariate Analysis." Academic Press, 1979.
- [26] J. Schmid Jr., "The Relationship between the Coefficient of Correlation and the Angle Included between Regression Lines". *The Journal of Educational Research* 41 (4), 1947.
- [27] Hexoskin data validation. Available at: [https://cdn.shopify.com/s/files/1/0284/7802/files/CSEP\\_Hexoskin\\_Poster\\_-\\_University\\_of\\_Waterloo.pdf?11148](https://cdn.shopify.com/s/files/1/0284/7802/files/CSEP_Hexoskin_Poster_-_University_of_Waterloo.pdf?11148)
- [28] A. Sheth, P. Anantharam, K. Thirunarayan, "kHealth: Proactive Personalized Actionable Information for Better Healthcare," *Workshop on Personal Data Analytics in the Internet of Things (PDA@IOT 2014), collocated at VLDB 2014*, Hangzhou, China, September 5th 2014.
- [29] C.R. Rao. "Estimation of variance and covariance components in linear models." *Journal of the American Statistical Association*, 67(337), 112-115, 1972.
- [30] P. Anantharam, T. Banerjee, A. Sheth, K. Thirunarayan, S. Marupudi, V. Sridharan, S. G. Forbis, "Knowledge-driven Personalized Contextual mHealth Service for Asthma Management in Children", *IEEE 4th International Conference on Mobile Services*, June 27 - July 2, 2015, New York, USA



# Conformance Testing of Healthcare Data Exchange Standards for EHR Certification

R. Snelick<sup>1</sup>

<sup>1</sup>National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA

**Abstract** – *Seamlessly sharing and using healthcare data as intended among distributed healthcare information systems is difficult. The adoption and adherence to clear and unambiguous standards can help manage this complexity. Well-defined standards, and conformance to those standards, provide the foundation for reliable, functioning, usable, and interoperable healthcare information systems. Recent federal programs (in the US) have included incentive payments for healthcare providers who adopt and “meaningfully use” certified electronic health record (EHR) technologies; however, unless these products are developed using clearly-defined standards, the adoption rate will increase, but the promise of improved quality of healthcare will not be realized. The proliferation of healthcare information systems designed without compliance to standards will likely exacerbate, not lessen, current patient care challenges by creating a landscape saturated with systems lacking usefulness, usability, and interoperability that will be rejected by the end-user community. Additionally, the standards must be used and deployed as intended, and conformance testing is the process that helps ensure adherence to the standards. In this paper, we explore conformance testing and the tools that are used to perform HL7 (Health Level Seven) v2-based conformance testing for certification of EHR technologies.*

**Keywords:** Conformance; Conformance Testing; Data Exchange Standards; Healthcare Information Systems; Interoperability.

## 1 Introduction

As described in [1], use of electronic health records (EHRs), especially systems with clinical decision support capabilities, has been shown to enable quality improvement in healthcare as well as to help reduce the cost of that care when used regularly in the practice of medicine. Recognition of these findings led to the enactment of the Health Information Technology for Economic and Clinical Health (HITECH) Act, which provides funding for incentive payments to physicians and hospitals that adopt health information technology (HIT). Initially focusing on adoption of EHRs, approximately \$17 billion in the Center of Medicare and Medicaid Services (CMS) incentive payments were made

available through CMS’s HITECH-based EHR Meaningful Use (MU) Program, to be paid to providers that attest to or demonstrate “meaningful use” of “certified” EHR technology (CEHRT) [1]. In-line with the CMS program, the Office of the National Coordinator (ONC) published EHR certification criteria [2] and established a program for certifying EHR technologies [2]. ONC, in collaboration with the National Institute of Standards and Technology (NIST), developed test procedures and conformance test tools [3] based on the ONC’s EHR certification criteria. EHR technologies are tested for compliance to the criteria by ONC-Accredited Testing Laboratories (ATL). This paper provides an overview of the testing approach and test tools used by the ATLs to verify that vendors’ EHR technologies meet the certification criteria that specify HL7 v2 data exchange standards.

## 2 HL7 V2 Data Exchange Standard

There are numerous healthcare data exchange standards in the US and internationally for communicating administrative and clinical data. The most widely used standard is the Health Level Seven (HL7) Version 2.x *Application Protocol for Electronic Data Exchange* (hereafter HL7) [4,5]. This standard is designed to support application-to-application message exchange. An HL7 *message* contains data for real-world events such as admitting a patient or sending a laboratory result (e.g., a CBC—complete blood count). For each event, HL7 defines an *abstract message definition* that is composed of a collection of segments (data units, e.g., Patient Demographics) in a predetermined sequence. Rules for building an abstract message definition are specified in the HL7 message framework, which is hierarchical in nature and consists of building blocks generically called *elements* [4,5]. These elements are *segment groups*, *segments*, *fields*, *components*, and *sub-components*. The requirements for a message are defined by the message definition and the constraints placed on each message element. The constraint mechanisms are defined by the HL7 conformance constructs which include usage, cardinality, value set, length, and data type. Additionally, explicit conformance statements are used to specify other requirements that can’t be addressed by the conformance constructs.

An HL7 conformance profile (also referred to as a message profile) is a constraint on the base standard that defines specified requirements for a given use case (event), a common practice in HL7 specifications [5,6]. The message profile can be represented as an XML document, which forms the basis for conformance testing. Figure 1 shows an example snippet (condensed) of a lab result profile [7]. Each element in the message profile is listed along with its associated attributes. For a more detailed description of a message profile refer to the HL7 standard [4] and [5,6]. It is important to note that the attributes and constraints a profile applies to a message provide a clear and unambiguous definition, thereby facilitating the design, implementation, and testing of interfaces [5,6]. The NIST EHR conformance test tools use the XML message profile as the basis for validation [3,5,8,9].

Fig. 1. Snippet from a Message Profile

```
<Segment Name="OBX" Max="1" Min="1" Usage="R"
LongName="Observation/Result">
  <Field Name="Observation Identifier" Max="1" Min="1"
Usage="R" Datatype="LRI_CWE_CR" Table="VS_LOINC">
    <Component Name="Identifier" Usage="R" MinLength="1"
MaxLength="20" Datatype="LRI_ST">
      <ConformanceStatement id="LN-001"><EnglishDescription>If
CWE.3 (Name of Coding System) is valued "LN" then CWE.1
SHALL be a valid LOINC code identifier format.
</EnglishDescription><Assertion><Custom id="1"
className="gov.nist.healthcare.mu.lri.custom.Loinc"/></Asse
rtion></ConformanceStatement>
    </Component>
    ...
  <Field Name="Observation Value" Max="1" Min="0" Usage="RE"
Datatype="varies"></Field>
  <Field Name="Abnormal Flags" Max="*" Min="0" Usage="RE"
Datatype="LRI_IS" Table="0078">
```

The message structure defines a *template* to which the message must comply; it explicitly defines the elements and the sequencing of the elements in a message instance. Conformance constructs are used to define and constrain requirements on message elements. Figure 1 provides a representative lab result observation that is defined by the OBX segment [7]. The OBX segment has multiple fields such as Observation Identifier (e.g., Cholesterol), Observation Value (e.g., 196), Units (e.g., mg/dL), and Abnormal Flags (e.g., N). There are other segments in the message definition such as PID—Patient Identification. Fields can also contain structure, i.e., components and sub-components. For every element, constraints are defined, e.g., Usage (indicates if the element is required, conditional, etc.), Cardinality (indicates the number of times the element may occur), or Value Set (indicates a defined vocabulary). The message structure and the element constraints define the requirements and are used for message validation.

The 2014 Edition/Stage 2 ONC EHR certification standards and criteria specify four HL7 v2 implementation guides that apply to five certification criteria: (1) transmission to immunization registries, (2) syndromic surveillance to public health agencies, (3) transmission of reportable lab results to public health agencies, (4) transmission of electronic lab results to ambulatory providers, and (5) incorporation of lab tests and results [2]. A conformance profile is defined for each interaction (event) covering the specific use, for example, sending an immunization record from the EHR system to the Immunization Information System (IIS). The conformance test tools described below are used to ensure that EHR systems correctly implement this interface standard (profile).

### 3 Conformance Testing

Conformance testing is a process that determines if an entity (message, document, application, system, etc.) adheres to the requirements stated in a specification. Conformance testing is a multi-faceted operation that can range from a simple assessment of the validity of a message value to a nuanced determination of a system's reaction to a complex sequence of events. Conformance testing strives to establish a degree of confidence in the conformity of a given entity (implementation) based on the quantity and the quality of the tests performed. Interoperability testing assesses whether applications (or software systems) can communicate with one another effectively and correctly, and whether they can provide the expected services in accordance with defined requirements (i.e., have a common understanding and use of the data exchanged). Such testing is critical, since many modern system architectures are designed as distributed systems and rely on seamless operations.

NIST developed the HL7 v2 validation tools for ONC 2014 Edition/Stage 2 HIT certification testing. These tools covered the standards and criteria described in section 2. Although, the initial focus of this effort targeted ONC HIT certification, the tools are equally applicable and valuable for use at site installations. In fact, a number of local public health registries have incorporated the tools into their operational environments or, as in the case of the Arkansas Department of Health for instance, require them for on-boarding [10].

### 4 Testing Sending Applications

When testing the ability of the System Under Test (SUT) to create messages<sup>1</sup>, the focus of the conformance testing is on validating the message produced by the sending

<sup>1</sup> The concepts apply equally to documents and other message protocols.

system (e.g., an EHR). The *sender* SUT is treated as a “black box” — only the content of the message is of interest not how the message is created or transmitted.

The NIST conformance test tools used for validating sending systems have two operational modes: (1) Context-free and (2) Context-based. The Context-free mode validates any HL7 v2-based message created by the SUT for the given subject (e.g., immunization messaging, lab result messaging). It is not dependent on a specific use case instance, Test Case, or specific test data content. The Context-based mode validates messages created by the SUT that are associated with a given use case instance and a Test Case that includes specific test data that are entered into the SUT. The validation assesses the technical requirements and content-specific requirements specified in the Test Case. Context-based validation expands the test space, enabling more comprehensive testing (e.g., testing of conformance usage constructs such as “conditional” and “required, but may be empty”).

Context-based conformance testing of the technical requirements and capabilities of the EHR technology is central to certification testing. Through collaboration with subject matter experts from the ONC Standards and Interoperability (S&I) Framework, the Centers for Disease Control and Prevention (CDC), the Association of Public Health Laboratories (APHL), and the International Society for Disease Surveillance (ISDS), NIST developed the Test Cases that targeted the most important use cases and capabilities specified in the referenced standards.

The test data are provided to assist the Tester in verifying that the vendors’ EHR technologies are capable of supporting the required functions. Verifying the ability to support the specific test data content is a secondary aspect to the certification testing. Testing and verification related to specific content usually are more appropriate for local installations of the EHR technologies; however, for certain aspects of a certification Test Case, examining exact content is necessary to verify that a capability exists in the EHR technology. An added benefit of providing realistic test data for common use cases is reinforcement of the expected interpretation and use of the referenced standards.

Both Context-free and Context-based modes are useful for message validation. Since Context-free testing is not tied directly to Test Case data, any message instance can be validated. This method suits site installations well, enabling in-house testing on messages that are tailored to local requirements. Context-based testing is driven by Test Cases, targeting specific Test Scenarios that enable more precise testing. Context-based testing is the method

used in the ONC HIT Certification Program, however, the NIST HL7 v2 conformance test tools [3,9] support both modes of validation.

#### 4.1 Case Study: Transmitting Lab Results

In this section, a case study based on the ONC Edition 2014 certification criterion for the transmission of laboratory results is used to explain the principles of Context-based testing for a sending application.

The focus of conformance testing for a sending system is on validating the message. The SUT is treated as a “black box”—how the message is created or transformed is not in scope. If we consider a laboratory information system (LIS) or a laboratory module that is integrated with an EHR system (hereafter called “lab component”), testing is not concerned with the detailed architecture of the lab component, but rather with what it produces (a lab results message) based on a given set of inputs (i.e., a lab results interface Test Case). The “black box” can consist of a self-contained lab system or multiple interrelated modules. The Use Case for the transmission of lab results could consist of the following steps:

1. A lab test is ordered for a patient
2. The specimen is collected (if applicable), and is received and processed by the lab
3. The lab result is produced, imported, and stored by the LIS
- 4. The lab result message is created**
5. The lab result is transmitted to an ambulatory electronic health record (EHR) system
- 6. The lab result is incorporated into the ambulatory EHR system**

The scope covered by the ONC transmission of lab results criterion is step 4 above – the lab result message is created. Step 6 – the lab result is incorporated into the ambulatory EHR – is covered by the incorporate lab results criterion; see section 5 for a case study for testing receiving systems.

Test Cases are provided for specific laboratory tests for which a lab results message will be imported into the conformance test tool (e.g., a CBC—Complete Blood Count). A Test Case consists primarily of a narrative Test Story (one possible path described by the Use Case) and a Test Data Specification. The Test Story describes a real world situation and provides the context for the Test Case. The Test Story also provides details associated with the Test Case such as pre-conditions, post-conditions, test objectives, and notes to testers. The Test Data Specification provides the data associated with the Test Story and consists of typically available information in the clinical setting. Together the Test Story and the Test Data Specification provide sufficient information to

be entered into the SUT for a particular Test Case, e.g., creating a lab results message. A Message Content Data Sheet is provided to show a conformant message instance for the Test Case. It also lists the category for each message element, indicating the kind of data and the expected source, which are based on the test case and test case objectives. How the data are categorized is directly related to how the message content is validated by the Test Tool. In some cases the validator is examining a message element for the presence or absence of data, and in other cases it is examining the message element for both the presence of data and exact content. The Message Content Data Sheet provides the evaluation criteria (expectations) and can be thought of as the “answer” to the “question” given in the Test Story and the Test Data Specification.

Fig. 2. Context-based Validation Test Flow

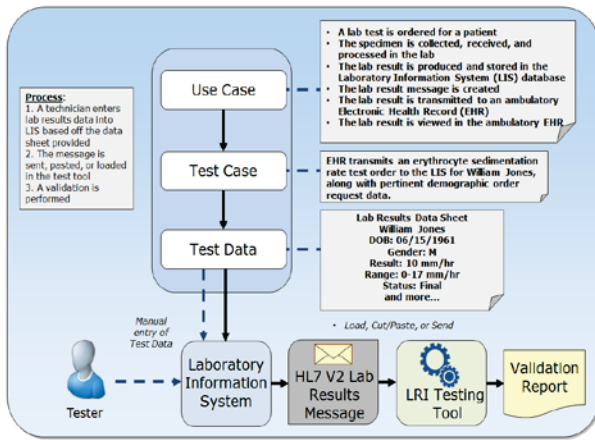
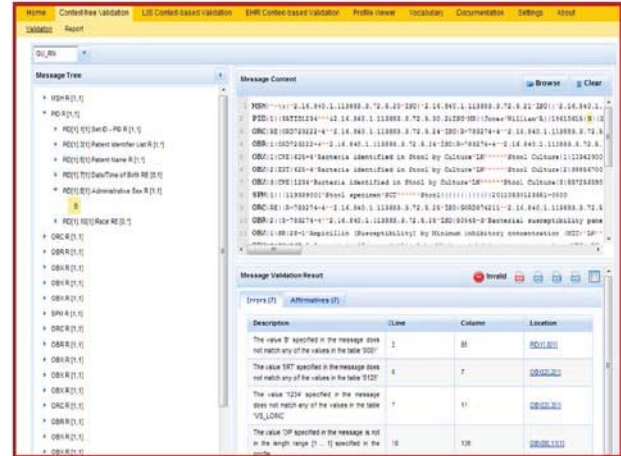


Figure 2 summarizes the Context-based testing flow for a representative Test Case. The Test Case artifacts are accessed by the Tester, the test data are loaded into the LIS, and a lab result message is generated. The Tester selects the corresponding Test Case in the Test Tool, imports the generated message, and the Test Tool validates the message based on the requirements in the Lab Results Interface (LRI) specification [7]. The web-based Test Tool provides interactive validation results for each message along with validation report documents.

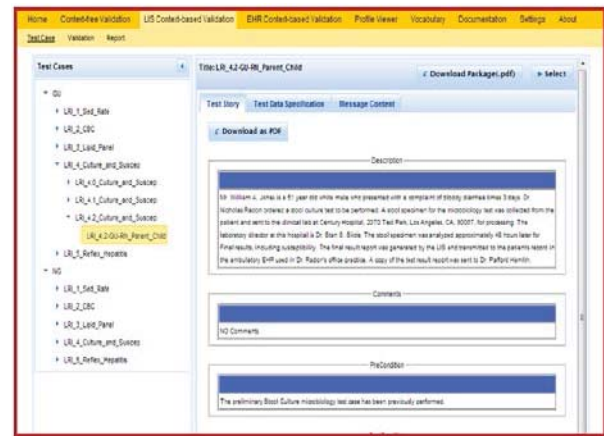
The Context-free and Context-based testing modes in the NIST Test Tools have certain basic features in common. When the user selects the Context-free tab, they then select the desired message profile and import the test message into the Test Tool. Once the message is imported, the message validation is performed automatically (Figure 3). The left-hand panel of the Validation screen shows the user a tree structure view of the message where individual data-content can be examined. The upper-right panel is the message content window. The validation results are displayed in the lower-right panel of the tool and include a description of the error and its location.

Fig. 3. Message Validation Panel



In the Context-based mode the user first selects a specific Test Case that provides a particular scenario and test data. These data are entered into the LIS which creates a message that corresponds to the test data. The validation process then proceeds much like the Context-free mode, except the validation is bound to specific data requirements defined by the Test Case. Figure 4 shows a screenshot of the Test Case panel that includes the Test Story (shown), the Test Data Specification, and the Message Content. The tool also includes a Profile Viewer and Vocabulary tab that allow browsing of the requirements specified in the implementation guide. The Test Cases and the Context-based validation are linked; i.e., in addition to validating the technical requirements specified in the implementation guide, the Test Tool performs selective content validation based on the provided Test Data Specification with the associated data categorization.

Fig. 4. Test Case Panel



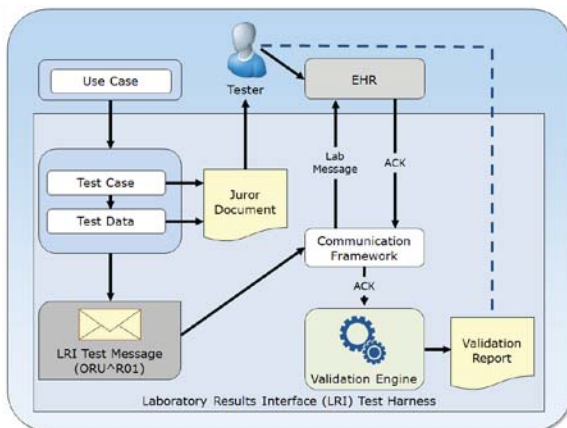
## 5 Testing Receiving Applications

Testing a *receiving* system is a challenge, because, typically, the system does not produce a *tangible* object (e.g., a message) for the Tester to assess directly. From

the available testing approaches used for testing receiving systems, the Inspection Testing approach has proven most suitable for the ONC HIT certification testing environment.

Inspection Testing relies on human validation (a visual inspection) of the SUT in order to collect evidence for the conformity assessment. Usually, the Inspection Testing process involves priming or knowing the state of the receiving system; providing a known and documented stimulus to the system; and evaluating the system's response to the stimulus against expected results based on the input and requirements.

**Fig. 5.** Testing Incorporation of Lab Results



A “test harness” contains the Test Case and associated test material, such as the test message and a Juror Document. The Juror Document (Test Case-specific inspection check list) is used by the Tester to inspect the receiving SUT for conformance to the specification. The information contained in the Juror Document is based on the data provided in the test message, the known state of the system, and requirements listed in the given test criterion. During the inspection process, evidence of the SUT's conformance can be obtained through a variety of methods, including viewing the system's display screens, browsing the system's data base, viewing the configuration files, or other mechanisms supported by the SUT.

### 5.1 Case Study: Lab Results Incorporation

Testing for the incorporation of laboratory results provides a good example of the challenges faced when testing a receiving system. No output artifact is produced that can be assessed directly by the Tester during this test [1]. For this ONC criterion, the ambulatory EHR, as the receiving SUT, is examined for evidence of the *incorporation* of laboratory results information from the received message and also for the ability to display seven types of information that are part of a laboratory results report (per requirements adopted from the Clinical

Laboratory Improvement Amendments (CLIA), which are regulatory standards for clinical laboratory operations in the US). The ONC criterion for incorporation of laboratory results specifies the Laboratory Results Interface (LRI) Implementation Guide [7] for generating the laboratory results message, and the related conformance testing involves a Juror Document and a human inspector. The content of the Juror Document is derived mostly from the Test Case and test message. Figure 5 illustrates the testing flow when using the NIST conformance test tool for the incorporation of laboratory results test procedure.

**Fig. 6.** Juror Document Panel



The Test Tool [9] provides a test harness that interacts with the EHR SUT, simulating the function of an LIS (or a laboratory component) that would create the LRI message. The use case described in the LRI implementation guide for creating lab result messages is the counterpart to the use case described for incorporating these messages; therefore, the same Test Cases developed for creating lab result messages can be used for incorporation of these messages, which allows for reuse of certain testing artifacts. The EHR SUT is primed with data (i.e., patient demographic information) to enable incorporation of the lab result message data elements into a specific patient's record; and the Test Tool, EHR, and test message are configured to enable communication between the systems.

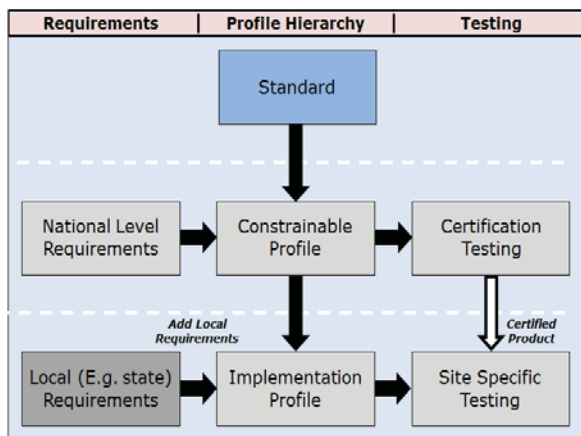
The lab result test message is sent from the LRI test harness to the EHR SUT, and the Juror Document is generated automatically by the test tool infrastructure based on the Test Case and test data. The Tester obtains the Juror Document and uses it to examine the EHR and to verify and document the presence or absence of the data elements transmitted in the test message. The data elements are categorized in the Juror Document according to how they are to be verified. For example, some data elements must be displayed to the clinical user on the EHR screen as well as stored in the EHR, while

other data elements are required to be stored or derivable only. The results gathered by the inspector are used in combination with the Validation Report from the ACK message to determine if the SUT passed or failed the test.

## 6 Perspectives of Testing

Profiling and the application of a profile hierarchy for specifying requirements of data exchange standards are critical for achieving interoperability [6]. A conformance profile is a refinement to either the underlying standard or another conformance profile, and it normally specifies constraints on messages or documents. A relationship can be drawn between the profile level and the type of testing that can be performed, as illustrated in Figure 7. The relationships shown are not the only relationships possible, but they are the typical ones. SDOs (Standards Development Organizations) that create implementation guides often do so at the national (realm) level. These constrainable profiles defined within an implementation guide express a minimum core set of capabilities that each implementer must meet. Beyond this core definition, a certain amount of flexibility is allowed for data elements not fully qualified in the implementation guide; for example, the “optional” usage defined for some elements could be redefined by implementers as “required”.

Fig. 7. Profile Level and Testing Relationship



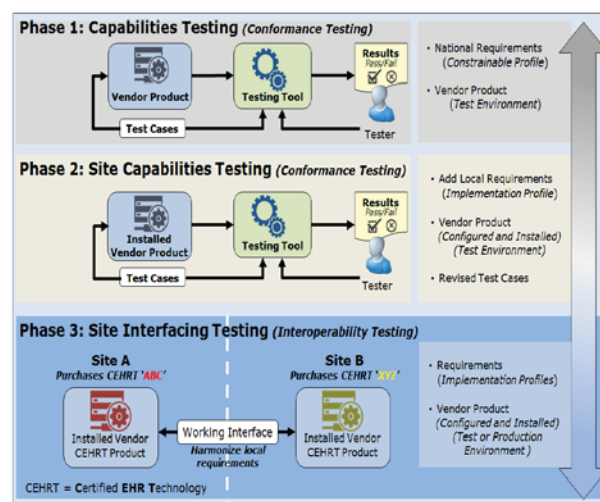
The established *baseline* ensures a consensus level of functionality that satisfies the targeted use cases. Vendor and local implementations further constrain and define requirements that are compliant with the national level requirements. Local, that is site<sup>2</sup> specific, variations are denoted explicitly.

National certification programs (such as the ONC HIT Certification Program in the US) develop and/or

<sup>2</sup> The term “site” in this context can mean a single site, multiple sites, or a group of sites that define the same (implementation) profile requirements.

reference national level profiles for their certification criteria. One of the objectives of certification testing is to assess whether the capabilities in each vendor’s product meet the requirements defined by the specified profile. Capability Testing is the type of testing used for ONC HIT certification testing and is based on constrainable profiles as depicted in Figure 8. A key point to bear in mind for Capability Testing is that its purpose is to verify that a product has the required capabilities, not to verify how the product might be used when installed in a production environment. Once any needed additional local requirements are established and documented, site testing is performed using an implementation profile (i.e., a completely defined specification). Site specific testing focuses on the ability of a product to support its intended use at an actual installation, which may be based on partner agreements.

Fig. 8. Levels of Testing



Certification testing seeks to ensure that every product that is certified supports the capabilities defined by the national level standard. National level certification testing brings a set of stakeholders one step (or phase) closer to achieving interoperability, but it is only the first step (see Figure 8). It is incorrect to assume that installing certified products will lead to “out of the box interoperability” when interfacing two or more of these products. The scope of ONC HIT certification testing is phase one. After Capability Testing of vendor products is performed in a test setting, a second round of Capability Testing that includes testing based on local requirements should be performed; we refer to this level of testing as Site Capability Testing (See Figure 8).

At the national standard profile level, local requirements and variations have not been taken into account. Once local agreements are defined and the profiles have been documented, site specific testing can occur. The distinction between the different profiles and the associated levels of testing is important. Capability

Testing occurs in phase one of the process and focuses on conformance testing. For site installations, the baseline requirements are customized to meet local requirements, and additional conformance testing needs to occur. This local conformance testing is required to ensure that the local requirements are implemented and that the national requirements have not been compromised (think of this as a form of “regression” testing). Once all parties participating in the site installation have completed this second round of conformance testing, then interoperability testing can proceed.

Step (or phase) three focuses on interoperability testing, and, ideally, conformance testing should continue to be included in the process. The need to include conformance testing here is especially critical if the implementations are being modified to achieve interoperability. It is important that conformance is not compromised to obtain interoperability. The sites wishing to interoperate likely have purchased certified products that have been customized to meet site requirements and have tested those implementations accordingly. Site Interface Testing is employed to determine that both data exchange and data use meet the business requirements. Such testing addresses the question: does the interface work for the intended use case? Site Interface Testing can be performed in a test or production environment.

Although we have presented the testing steps as a group, the concepts of conformance and interoperability testing are orthogonal. Conformance testing is performed on the various profile levels in the hierarchy, and the product is tested in isolation. Interoperability testing is performed among a set of products, be it in a test environment (such as the IHE Connect-a-thons [11]) or at a production site. Although orthogonal in nature, the sequence in which testing should occur is progressive and must take into consideration the realities of the production setting in which the HIT technologies are to be used. There is limited value in performing interoperability testing without prior agreements and conformance testing.

## 7 Summary

Improved outcomes, clinical decision support, and patient safety are a few of the many benefits provided by interoperable healthcare information systems such as EHRs [12]. To achieve interoperable systems, products must be developed to a set of well-defined standards that are universally adopted. To help ensure that the standards are implemented correctly, conformance testing is necessary. In the US, ONC has established a program to certify EHR technologies that uses NIST conformance testing tools. Certified products ensure a level of capabilities, which is the critical first step towards achieving interoperable systems. Beyond this step, refinement of standards within the framework of the

established base standard is often necessary to accommodate site specific requirements. Subsequent conformance and interoperability testing also is necessary. Following this course of action will drive the industry closer to the goal of interoperable healthcare information technologies.

## 8 References

- [1] *Understanding Meaningful Use with a Focus on Testing the HL7 V2 Messaging Standards*. R. Snelick, S. Taylor, HL7 May 2013 Newsletter. <http://hl7.org>.
- [2] ONC Certification Programs; <http://www.healthit.gov/providers-professionals>
- [3] *NIST Resources and Tools in Support of Health IT Standards and the ONC Meaningful Use certification program*. <http://healthcare.nist.gov/>
- [4] Health Level 7 (HL7) Standard Version 2.7, ANSI/HL7, January, 2011, <http://www.hl7.org>.
- [5] *Towards Interoperable Healthcare Information Systems: The HL7 Conformance Profile Approach*. R. Snelick, P. Rontey, L. Gebase, L. Carnahan. Enterprise Interoperability II: New Challenges and Approaches. Springer-Verlag, London Limited 2007 pp. 659-670.
- [6] *Principles for Profiling Healthcare Data Communication Standards*. R. Snelick, F. Oemig. 2013 Software Engineering Research and Practice (SERP13), WORLDCOMP'13 July 22-25, 2013, Las Vegas, NV.
- [7] *HL7 Version 2.5.1 Laboratory Results Interface (LRI) Implementation Guide*. Draft Standard for Trial Use. July 2012. <http://www.hl7.org>.
- [8] *A Framework for testing Distributed Healthcare Applications*. R. Snelick, L. Gebase, G. O'Brien. 2009 Software Engineering Research and Practice (SERP09), WORLDCOMP'09 July 13-16, 2009, Las Vegas, NV.
- [9] *NIST Laboratory Results Interface (LRI) EHR Meaningful Use Conformance Testing Tool*; <http://hl7v2-lab-testing.nist.gov/mu-lab/>
- [10] Arkansas Department of Health Onboarding. <http://www.healthy.arkansas.gov/programsServices/MeaningfulUse/Immunizations/Pages/OnBoarding.aspx>
- [11] *IHE Connect-a-thon* <http://www.ihe.net/Connectathon>
- [12] Benefits of Electronic Health Records (EHRs) <http://www.healthit.gov/providers-professionals/benefits-electronic-health-records-ehrs>

# Context-Mediated Semantic Alignment of Heterogeneous Electronic Medical Record Systems For Integrated Healthcare Delivery

Guoray Cai

College of Information Sciences and Technology  
Penn State University, University Park, PA 16802  
Email: cai@ist.psu.edu

**Abstract**— Healthcare organizations are facing the challenge of integrating their electronic medical records (EMR) for the purpose of unified healthcare service delivery. Through an analysis of the nature of medical data and the use of medical records, we argue that the most significant barrier for service integration is the semantic heterogeneity across fragmented and disintegrated healthcare systems and health records. Based on our understanding of the requirements for EMR semantic integration and the limitations of the existing solutions, we present an approach for modeling the context-dependent nature of medical data semantics. The approach is formalized into a framework (*Context-Mediated Semantics Interoperability*), which serves as the basis for computational support to medical service integration. Our initial contextualization effort focused on capturing the heterogeneity across different types of specialties, as well as the semantic differentiation for individuals, organizations, communities, and social use.

**Keywords**—service integration; healthcare delivery; electronic medical records; contextualization; ontology; semantic web;

## I. INTRODUCTION

Modern medical service enterprises must deal with integration of information nuggets and silos that exist in different units of health services and medical operations in order to enable meaningful sharing, reuse, and collaboration [3-7]. A huge challenge towards that goal is the semantic barrier of communication and interoperability for diverse use and reuse of heterogeneous medical data [1, 2, 8, 9]. While differences in language, data schemas, data models, and communication protocols can be reconciled through existing solutions (such as standardization, model integration, ontology integration) [10-12], these methods do not work well when semantic heterogeneity presents.

This research seeks deeper understanding of the nature of semantic heterogeneity in electronic medical data and service integration, and proposes a conceptual framework towards meaningful solutions targeting healthcare domain. The US Congress made it to law (HITECH - *Health Information Technology for Economic and Clinical Health*) Act to provide financial incentives to physicians and other healthcare stakeholders to adopt Electronic Health Records. A core requirement of the Act is to demonstrate “Meaningful Use”[13-16]. One of the foremost barriers is the fragmented and disintegrated healthcare systems and

health records. With as many as 20 healthcare providers per patient, a patient’s health record may include a diverse set of views and interpretations of the patient’s health conditions [17]. They often present inconsistent and even conflicting records of diagnosis, treatment plans, and instructions. Effective communication and collaboration among providers is critical to achieve quality and safety outcomes [17, 18]. A recent national analysis of the value of interoperability suggested that fully interoperable healthcare systems could save the nation \$77.8 billion annually[19]. Thus, it is a high priority to implement interoperable EMRs across healthcare organizations. Pure ontology-based approaches for interoperability (such as OpenEHR project [5]) have made limited success. This is not surprising, based on the lessons learned from other domains.

As the first step towards a solution, this paper presents a critical review of the literature to establish a conceptual framework for EHR semantic integration. The review is driven by the following questions: (1) *what is the nature of ‘medical work’ and what are the roles of the (electronic) medical record?* (2) *what are the sources and causes of semantic heterogeneity?* (3) *what are the requirements for HER semantic integration?* (4) *what are the solutions available and how well they work?* We seek answers from a diverse literature in the fields of health informatics, healthcare delivery, information science, and semantic web.

The second part of this paper proposes a framework for addressing the semantic heterogeneity problem in the integration of electronic medical records. The core idea is the extension of existing ontology-based semantics data models to include the contexts of use as a new layer of semantics in the modeling stack. We will describe the framework for semantic interoperability, which is called *Context-Mediated Semantics Interoperability* (CMSI framework). It specifies a semantic model in terms of three related components: activity-centric context representation [20, 21], contextualized ontology space [2], and context mediated semantic exchange [22]. Refinement of the framework as well as the computational methods will be done through a set of use scenarios taken from two types: heterogeneity across different types of specialties, (2) semantic differentiation for individuals, organizations, communities, and social use.



## II. ENTERPRISE INTEGRATION IN HEALTHCARE ORGANIZATIONS

American healthcare system is facing unprecedented challenges and opportunities, and is due for major re-engineering of the whole system. On one hand, health care is substantially underperforming on most dimensions: effectiveness, appropriateness, safety, cost, efficiency, and value. According to recent reports [17, 23],

- U.S. healthcare system is the most expensive healthcare system in the world. It yet it is among the least effective. US healthcare spending was about \$7,439 per person and accounted for 16.3 percent of the nation's gross domestic product (GDP) in 2007 and will trend upward reaching 19.5 percent of GDP in 2017 [23].
- It contributes to only 5 to 10 percent of total health—an estimated 3.5 to 7 years of lifespan—derives from the health care delivery system
- Between 44,000 and 98,000 people die each year from preventable injuries sustained as part of care delivery in U.S. hospitals.
- Waste: According to the National Academies, between 30 and 40 percent of healthcare costs— more than half a trillion dollars per year— is spent on “ overuse, underuse, misuse, duplication, system failures, and unnecessary repetition, poor communication, and inefficiency.”

A fundamental cause of these problems is the fragmented nature of the healthcare delivery system. Fragmentation exists in all levels of healthcare system, from organizational, incentive structures, service provisions, and information technology infrastructure [24]. In particular, there is a high degree of fragmentation in treatment and care. A patient in the US is served by as many as 20 healthcare providers in lifetime. A patient's health record may include a diverse set of views and interpretations of the patient's health conditions[17]. They often present inconsistent and even conflicting records of diagnosis, treatment plans, and instructions. Fragmentation of treatment and care is the result of increasing specialization in health care that has led to increasing regimens by specialists. Unfortunately, the holistic focus and interdependence of treatment nuggets tends to be lost.

In a fragmented healthcare delivery system, there is ever stronger the need of integration and synthesis. High-quality care requires the smooth flow of information across diverse providers working within various organizations in both in-patient and outpatient settings. A first step in fixing the system would be to build the interoperable electronic medical records that should improve coordination among providers and reduce gaps in care. A fully realized interoperable healthcare IT system could reduce errors, improve communication, help eliminate redundancy, and provide numerous other benefits that would protect patients and save up to tens of billions of dollars per year.

The central challenge to achieving such a system is interoperability—the ability of data systems, medical devices and software from different vendors based on a diverse array of platforms to share patient EHRs, electronic physician orders for lab tests and drug prescriptions, electronic referrals to specialists, electronic access to information about current recommended treatments and

research findings, and other information. The vision, as stated by the “healthcare information enterprise *integration* initiative” [4] is to transform the healthcare system from isolated treatment episodes towards one that is *patient-centric, coordinated, and continuous treatment process* involving multiple healthcare professionals and various institutions. A recent national analysis of the value of interoperability suggested that fully interoperable healthcare systems could save the nation \$77.8 billion annually [19]. Thus, it is a high priority to implement interoperable EHRs across healthcare organizations. President Obama initiated a massive national HIT program and allocated \$20 billion in funds for Health Information Technology for Economic and Clinical Health or HITECH, etc.

## III. THE NATURE OF ‘MEDICAL WORK’ AND THE ROLES OF ELECTRONIC HEALTH RECORDS

Medical work is the combination of diagnosis, treatment, and prevention of disease, illness, injury, and other physical and mental impairments in humans. It involves caregivers who practice medicine, chiropractic, dentistry, nursing, pharmacy, allied health, and other services. For optimal patient care, the various provider organizations and health professionals have to cooperate closely during patient care, often called shared care or integrated care. *Shared care* is defined as the continuous patient-oriented cooperation of hospitals, general practitioners (GPs), specialists and other health care professionals during patient care [5]. Shared care imposes great challenges on the availability and processing of information including trusting shared information, the correct and clinically safe interpretation of the information.

Electronic Health Records (EHRs) mainly involve clinical patient data including the personal and family history, the clinical state, the dispensed therapies, and other relevant information about the reached diagnostics and outcomes [25]. An EHR is used primarily for purposes of setting objectives and planning patient care, documenting the delivery of care and assessing the outcomes of care. It also includes information regarding patient needs during episodes of care provided by different health care professionals. This includes items like handwritten, typed, or electronic clinical notes; notes recorded from telephone conversations; all correspondence including letters to and from other health care professionals, insurers, patients, family, and others; laboratory reports; radiographs and other imaging records; electrocardiograms and printouts from monitoring equipment; audiovisuals; and other computerized/electronic records, including e-mail messages.

Since the first conception of electronic health records (HER) in the 1990s, the content, structure, and technology of EHR have been evolved, driven by the basic idea of supporting and enhancing health care, and improving service delivery and its quality [26]. In 2009, the American Recovery and Reinvestment Act mandated that hospitals in the USA move to electronic medical records (EMR) systems by 2014. It prescribes the implemented EMR to have the following characteristics: (1) *patient-centered*; (2) *longitudinal*-it is a long-term record of care, (3) *comprehensive* - it includes a record of care events from all types of care givers, providers and institutions tending to a

patient, not just one specialty, and (4) *prospective* - not only are previous events recorded, but also instructions and prospective information such as plans, goals, orders and evaluations.

An interoperable EMR with the above characteristics will serve as the foundation for translating the vision of evidence-based practice [27] into reality. An integrated EMR system serve as a record of the longitudinal health history of each patient is required to improve quality of care.

#### IV. THE SEMANTIC HETEROGENEITY OF MEDICAL RECORDS

Building interoperable electronic medical records must start with addressing data heterogeneity among systems and data collections. The proliferation of data heterogeneity among autonomous systems creates islands or silos of medical records, and presents major barriers towards full realization of the financial, clinical, and efficiency benefits conferred by HIT [28].

Crucial dimensions of data heterogeneity are syntactic, structural, and semantic heterogeneities (see Figure 1).

- (1) *Syntactic heterogeneity*: each system may represent medical data and knowledge using different encoding and communication format semantic interoperability paradigms, such as relational or object oriented models. Currently, there is no single universally accepted clinical data model that will be adhered to by all [29]. Syntactic interoperability ensures that clinicians can always send information to another provider and receive information which they can read.
- (2) *Schematic heterogeneity*: this refers to lack of interoperability due to the differences in the way each system structure objects and their relationships in medical contents. For example, objects in one system are considered as properties in another, or object classes can have different aggregation or generalization hierarchies, although they might describe the same Real World facts.
- (3) *Semantic heterogeneity*: Semantics is defined as the meanings of terms and expressions. Hence, semantic interoperability is the ability of information systems to exchange information on the basis of shared, pre-established and negotiated meanings of terms and expressions [5]. In semantically interoperable EMR, different information systems used by the various health care providers of shared care can understand the context and meaning of information provided by other systems. However, there are great variations across information systems in the descriptions of diseases, causes, and treatments, due to the highly specialized and dynamic service landscape.

Today powerful integration tools (e.g. application

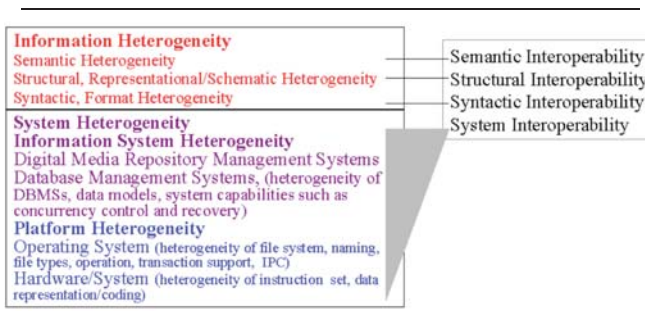


Figure 1. Levels of heterogeneity (after [1])

servers, object brokers, different kinds of message-oriented middleware, schema mapping, and workflow management systems) are available to overcome syntactical and schematic heterogeneity of medical record systems. Yet, *semantic heterogeneity* remains as a major barrier to seamless integration of patients' medical records. We will focus on semantic heterogeneity in the rest of this report.

A major contributor to semantic heterogeneity is the high degree of specialized service providers that apply different subset of medical knowledge to patients. Efforts for resolving semantic heterogeneity have taken multiple approaches. The most dominant approach is to establish standards for establishing meanings of medical terminologies and expressions. The development of vocabulary standards to support use of EHRs has been quite successful in the last two decades, mounted by the United States National Library of Medicine in its UMLS project, the UK National Health Service and its Centre for Coding and Classification, by SNOMED-CT International, and by the GALEN program of the European Community. Despite many years of concentrated and coordinated effort to build comprehensive medical terminology standards, a single agreed-upon system of global electronic medical reference terminologies and ontologies does not yet exist by today and there has been doubts if such a goal is possible [2, 30].

While a single global medical ontology does not exist, there are many vocabulary systems that are in use including anatomy concepts, ICD9, ICD-O-3, INDEX VIRUM, LOINC, MEDCIN, Med- DRA, SNOMED and the clinical drug codes being developed by the Veterans' Administration and the Food and Drug Administration. In addition, the National Library of Medicine defined global electronic medical ontologies [26, 31-33].

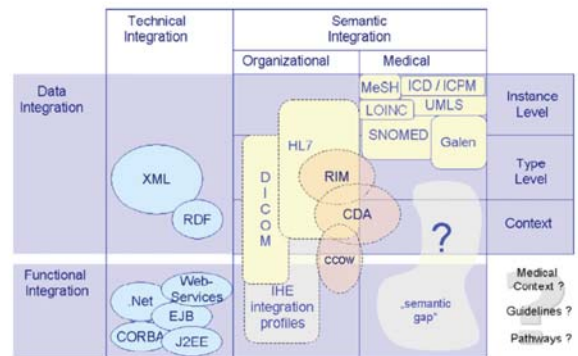


Figure 2. Contribution of different standards to application integration [6]

Existing terminology systems that are developed in academic research projects are fundamentally flawed from the point of view of practical use in scalable systems, and this explains why commercial vendors of HIT systems rarely choose them. Figure 2 shows the gap that current standardization efforts have not addressed.

#### V. SEMANTIC HETEROGENEITY OF CLINICAL TERMINOLOGIES

Clinical terminology concerns the meaning, expression, and use of concepts in statements in the medical record or other clinical information system. Healthcare professionals manage constantly with minor differences of meaning and

even misunderstandings of information written by their colleagues or received from other institutions. The ideal solution would be to reach a full consensus on a terminology standard that supports both the practical use by clinicians and by computer systems for automated processing and analysis. This seemingly straightforward problem (from semantic web perspective) has been partially proven to be extremely hard in medical information, for many good reasons [32, 34]. This is one of those application domains where “*special purpose solutions to small scale problems are easy, while fully scalable general purpose solution is extremely difficult.*” Another domain of similar nature is geographic information science which has been articulated by Cai [35]. Figure 3 shows the types of vocabularies that have been developed so far.

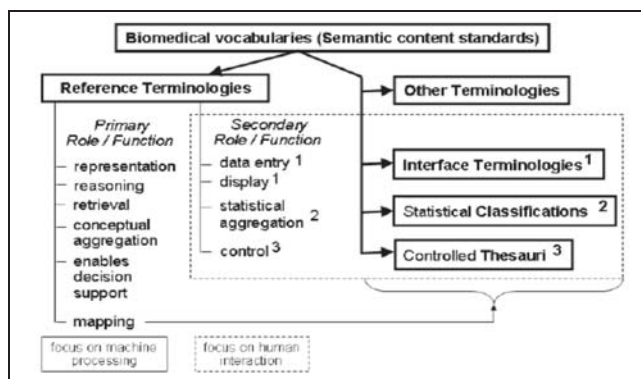


Figure 3 Types of biomedical vocabularies (after [2])

Reflecting on the slow progress in developing a comprehensive re-usable terminology for patient-centered systems, Rector [30] put forward ten reasons why clinical terminology for coordinated EMR is hard. I summarize a few of important ones below, since they are fundamental to the arguments I make in this paper.

**[Reason 1]** *It is difficult to scale up terminology solutions to the vast and the multiplicity of activities, tasks and users to be served by EMR.* The scope of medical knowledge across all specialties, for detailed clinical care, in particular, is orders of magnitude larger than the terminology needed to report simple diagnostic registers in a single specialty or even general practice. Scaling up by an order of magnitude or more is notoriously difficult, because the complexity of digital systems tends to increase exponentially with the scale of the vocabulary.

**[Reason 2]** *It is difficult to come up with terminology solutions that meet the requirements of usability by both human and machine processing.* Humans and machines process information very differently. Human are *compliant, flexible, tolerant, while machines require us to be rigid, fixed, and intolerant.* They present fundamentally conflicting requirements that we have no good solutions so far [36].

**[Reason 3]** *It is difficult to come up with terminology solutions when the complexity of clinical pragmatics is introduced.* Clinical pragmatics includes three aspects: clinical conventions, clinical expectation, and operational meaning. These pragmatic complexities often defy any

attempts to capture them fully in formal representations. Phrases do not literally mean what they say. Same phrases can be interpreted in multiple ways, with some of them more usual than others (for instance ‘Heart Valve’ would seem to mean “valve in the heart”, but physicians may understand it in four different ways, so additional information on common usage beyond either linguistics or the formal concept representation is required). The meaning of a term that is intended to cover may vary in scope with the situation within which an operational record was created, and is often undetermined outside that operational context. On the other hand, it is almost a necessity to embrace clinical pragmatics. A terminology that has no concern and proven relevance clinical pragmatics is useless. Unfortunately, developers of medical terminology have largely ignored the need to explicitly capture the pragmatic knowledge together with the terminology.

Given the above understanding of the requirements for medical terminology, the current solutions are seriously flawed! According to Rector [30] and Cimino [37], academic research has assumed that we can clearly separate linguistic knowledge, the medical concept systems, and pragmatic knowledge associated with clinical terminology.

**[Hypothesis of Separability]:** *For a clinical terminology, the representation of concepts and the relations between them can and should be separated from the linguistic knowledge about how these concepts are expressed in language and the pragmatic knowledge concerning how these concepts are used in dialogues with clinical users.*

Therefore, the whole problem has been cut into three separate sub-problems, each corresponding to a discipline and task:

1. *Clinical computational linguistics* – getting the language right
2. *Logical concept representation* – formal representation of concepts in ways which give rise to correct identification, classification, and retrieval of information in formal (computer) systems.
3. *Clinical pragmatics* – organizing information in ways expected by healthcare professionals and in ways that facilitate their daily work.

Because each of the three disciplines has developed separately and each uses different tools and techniques that are based on fundamentally different underlying principles, integrating such systems directly is difficult, perhaps impossible [18, 38]. The hypothesis of separability seriously underestimated the difficulties of integrating the three aspects of research outcomes in actual clinical information systems. One of the major difficulties in medical terminology has been the confusion of concepts and the words used to express those concepts. In particular, there exists ambiguity of mapping between the linguistic expressions and concept components in which one linguistic expression can be interpreted as more than one internal concept. Incorporating such ambiguity into formalisms and ontologies for clinical concept representation are hard.

## VI. OVERVIEW OF THE FRAMEWORK

In order to address the deficiencies of the existing approaches semantic heterogeneity of medical information

in a coordinated care enterprise environment, we proposed a framework, called *Context-Mediated Semantics Interoperability* (CMSI framework). The idea behind our framework is as follow:

- (1) Contexts should be explicitly represented; contextual knowledge should be associated with context representations; and contextual knowledge should guide all facets of an agent's behavior.
- (2) A theory of medical data semantics should include multiple ontologies ranging from top-level generic ontology to application specific ontologies. Each ontology is associated with a context that 'wraps' around it. Semantics of a specific ontology is local to its context. Ontologies are related through the generalization/specialization relationship of their contexts, as well as through explicit 'lifting' rules.
- (3) With contextualization of data semantics, it is no longer required for two communicating agents (or data sources) to have common ontological commitment. Instead, we rely on context alignment and shared contextual knowledge to constrain semantic interpretation.
- (4) Contextualization hides the heterogeneity of data at the ontology level, just like ontologies effectively hide the heterogeneity of data at the syntax level.
- (5) Contexts and ontologies are two semantic coordination mechanisms for interoperability, with contexts taking priority over ontologies. In other words, commonality in contexts can over-ride heterogeneity in ontologies, but not *vice versa*.

Our CMSI framework is the integration and extension of the work on *contextual ontology* (C-OWL) by Bouquet [39] and the work on *context schema* (C-schema) by Turner [40]. It consists of the following six components:

$\hat{C}$ : a context space which is a set of contexts  $\{C_i \mid i=1, \dots, N\}$ , where  $N$  is the total number of contexts

$\hat{O}$ : an ontology space which is a family of ontologies  $\{O_i \mid i=1, \dots, N\}$

$\hat{\Phi}$ : a set of inter-ontology bridging rules  $\{\Phi_{i,j} \mid (i, j \in \{1, \dots, N\}) \text{ and } (i \neq j)\}$ . Each  $\Phi_{i,j}$  is a set of rules that specify how elements of ontology  $O_i$  relates to elements in ontology  $O_j$ , if any relationship exists.

$\hat{\Psi}$ : a set of inter-context bridging rules  $\{\Psi_{i,j} \mid (i, j \in \{1, \dots, N\}) \text{ and } (i \neq j)\}$  that specify how context  $C_i$  relates to context  $C_j$ , if any relationship exists.

$\hat{\Theta}$ : a set of rules governing context coordination.

$\hat{\Omega}$ : a set of rules governing ontology coordination.

Each of these components is again a complicated structure. For more details, see [35].

## VII. POTENTIAL BENEFITS OF THE FRAMEWORK

The fundamental advantage of the framework originates from its premise that linguistic, conceptual and pragmatic issues of medical information must be considered together due to their close coupling in clinical interpretation and reasoning. In other words, we do not make any assumptions of separability as did in previous work (see section V). Instead, we make medical pragmatics explicit and use them

as formal semantic coordination mechanisms that embrace semantic heterogeneity across different use scenarios and clinical practice specialties. This perspective can be stated into two assumptions below:

1. First, data are always produced with a given purpose. Medical data complexity and specificity is directly tailored to that purpose [41]. Medical information is entangled with its context of production in that the meaning, hardness and significance of a piece of information cannot be detached from the specific purpose that structured the gathering of that information.
2. Objects in medical data mutually elaborate each other, rather than being isolated "atoms." For example, in the course of a patient's illness trajectory, data items are constantly reinterpreted and reconstructed.

Modeling and representing context can lead to several benefits [42, 43]:

*Economy of representation:* contexts can act as a focusing mechanism

*Economy of reasoning:* reasoning can be performed with the context associated with an information source (instead of the whole data)

*Managing inconsistent information:* As long as information is consistent within the context of the query of the user, inconsistency in information from different databases may be allowed

*Flexible semantics:* An important consequence of associating abstractions or mappings with context is that the same two objects can be related to each other differently in two different contexts. Two objects might be semantically closer to each other in one context as compared to the other.

Our solution bares some similarity with the OpenEHR initiative (<http://www.openEHR.org>) [5, 25, 44]. Our context schema can be viewed as an extension of Archetypes in OpenEHR specification. *Archetypes* are agreed models of clinical or other domain-specific concepts. From a technical point of view archetypes are formal specifications of clinical content. From a clinical point of view, archetypes serve an intuitive means to define and discuss and present clinical content [5]. It empowers the health professionals to define and alter the accurate knowledge and information they need in the granularity they need.

We summarize specific benefits of CMSI framework to semantic integration of EMR systems:

(1) *Supporting multiple points of view, poly-hierarchies and multiple levels of granularities.* This is in contrast to traditional ontology and terminology approaches that have limited applicability to a single use, single granularity, a single point of view, a mono-hierarchy and often to a single area of medicine. Unfortunately, reusable medical records require that terminologies support multiple points of view, poly-hierarchies and multiple levels of granularities.

(2) *Open-Endedness:* Not only is medicine big, it is open-ended. There are constantly new discoveries in medical knowledge. The framework is design to manage changes of language with respect to the underlying concepts, clinical practice, and the underlying system of concepts itself.

(3) *Minimize the difficulties of achieving consensus.* Achieving clinical consensus on existing terminologies has proved particularly difficult. Physicians disagree. Nurses disagree. Healthcare professionals disagree. Our approach supports flexible levels of consensus appropriate in each area and what areas can be left for local choice.

## VII. CONCLUSIONS

Following the proposed CMSI framework, a number of research questions must be fully answered in order to derive workable solutions for medical information service integration.

**[Question 1]** *What are the appropriate attributes that define a context schema?* We will incorporate the findings in context models literature [22, 45-52] and evaluate them using the insight we have achieved in earlier sections of this report to make proper choices. One important direction we will take is the activity-based representation of contexts, which has been explored before [21, 53]

**[Question 2]** *What are the appropriate structure and level of granularity of context schemas?* This is a central question that has to be answered based on deep understanding of the practical use of medical records, analytical reasoning of clinicians, and the structure of medical knowledge. I hope to develop a series of field studies observing and interviewing physicians in their workplaces.

**[Question 3]** *What are the reasoning capabilities across-contexts that are needed by medical record applications?* Context reasoning is important for real-time mediation of semantic exchange and negotiating meanings when heterogeneity exists. We will incorporate modal logic reasoning approach [51, 54] to address the complexity of reasoning on contexts.

Answering above questions require close interaction with healthcare delivery organizations (hospitals and patient care professionals) and observation of their practice use. We are in the process of developing concrete case studies and use scenarios through collaborations with local hospitals. These will be used as design artifacts for creating context schemas and alignment mechanisms.

## ACKNOWLEDGEMENT

Funding of this research is partially supported by a grant from the Center for Enterprise Architecture Research, Penn State University.

## REFERENCES

- [1] R. M. Colomb, "Impact of Semantic Heterogeneity on Federating Databases," *The Computer Journal*, vol. 40, pp. 235-244, January 1, 1997.
- [2] J. Ingenerf and S. J. Poepl, "Biomedical Vocabularies - the Demand for Differentiation," in *Medinfo 2007: Proceedings of the 12th World Congress on Health*. vol. 129, 2007, pp. 610-615.
- [3] G. Keith, "Integrating the healthcare enterprise (IHE)," *Synergy*, p. 15, 2003.
- [4] B. Gordon, "Healthcare Information Technology Enterprise Integration Act: report ", by United States. Congress. House. Committee on Science and Technology, 2007.
- [5] S. Garde, P. Knaup, E. Hovenga, and S. Heard, "Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes," *Methods of Information in Medicine*, vol. 46, pp. 332-343, 2007.
- [6] R. Lenz, M. Beyer, and K. A. Kuhn, "Semantic integration in healthcare networks," *International Journal of Medical Informatics*, vol. 76, pp. 201-207, 2007.
- [7] W. He and L. D. Xu, "Integration of Distributed Enterprise Applications: A Survey," *IEEE Transactions on Industrial Informatics*, vol. 10, pp. 35-42, 2014.
- [8] J. S. Williams, "Achieving interoperability: what's happening out there?," *Biomedical instrumentation & technology / Association for the Advancement of Medical Instrumentation*, vol. 46, p. 14, 2012.
- [9] S. L. Grimes, "The challenge of integrating the healthcare enterprise," *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society*, vol. 24, pp. 122, 124-124, 2005.
- [10] J. Park and S. Ram, "Information systems interoperability: What lies beneath?," *ACM Transactions on Information Systems*, vol. 22, pp. 595 - 632, 2004.
- [11] H. Hatanaka, "The Current State and Issues of Life Science Database Integration," *Journal of Information Processing and Management*, vol. 55, pp. 347-353, 2012.
- [12] M. S. Fetter, "Interoperability--making information systems work together," *Issues in mental health nursing*, vol. 30, p. 470, 2009.
- [13] K. Redstone, "Teaming up to make meaningful use of electronic health records," *Canadian Medical Association Journal*, vol. 184, pp. E131-E132, Feb 2012.
- [14] B. Tai, M. Boyle, U. Ghitza, R. M. Kaplan, H. W. Clark, and K. Gersing, "Meaningful Use of Electronic Behavioral Health Data in Primary Health Care," *Science Translational Medicine*, vol. 4, Feb 2012.
- [15] A. Siddiqui, K. J. Dreyer, and S. Gupta, "Meaningful Use: A Call to Arms," *Academic Radiology*, vol. 19, pp. 221-228, Feb 2012.
- [16] S. R. Patel, "Making Meaningful Use of Electronic Health Data," *Journal of Clinical Sleep Medicine*, vol. 8, pp. 19-20, 2012.
- [17] C. Grossmann, W. A. Goolsby, and L. A. Olsen, *Engineering a Learning Healthcare System: A Look at the Future: Workshop Summary*. Washington (DC): National Academies Press, 2011.
- [18] D. F. Lobach and D. E. Detmer, "Research Challenges for Electronic Health Records," *American Journal of Preventive Medicine*, vol. 32, pp. S104-S111, 2007.
- [19] N. A. Zhivan and M. L. Diana, "US hospital efficiency and adoption of health information technology," *Health Care Management Science*, vol. 15, pp. 37-47, Mar 2012.
- [20] J. E. Bardram, "Activity-based computing: support for mobility and collaboration in ubiquitous computing," *Personal and Ubiquitous Computing*, vol. 9, pp. 312-322, 2005/09// 2005.
- [21] J. E. Bardram, "Activity-based computing for medical work in hospitals," *ACM Trans. Comput.-Hum. Interact.*, vol. 16, pp. 1-36, 2009.

- [22] B. Prados-Suárez, C. Molinab, C. P. Yañezc, and M. P. d. Reyes, "Improving electronic health records retrieval using contexts," *Expert Systems with Applications*, vol. 39, pp. 8522-8536, 2012.
- [23] S. Kumar, N. S. Ghildayal, and R. N. Shah, "Examining quality and efficiency of the US healthcare system," *International Journal of Health Care Quality Assurance*, vol. 24, pp. 366-388, 2011.
- [24] R. D. Cebul, J. B. Rebitzer, L. J. Taylor, and M. E. Votruba, "Organizational Fragmentation and Care Quality in the U.S. Healthcare System," *The Journal of Economic Perspectives*, vol. 22, pp. 93-113, 2008.
- [25] M. Meizoso García, J. L. Iglesias Allones, D. Martínez Hernández, and M. J. Taboada Iglesias, "Semantic similarity-based alignment between clinical archetypes and SNOMED CT: An application to observations," *International Journal of Medical Informatics*, 2012.
- [26] G. J. Esper, O. Drogan, W. S. Henderson, A. Becker, O. Avitzur, and D. B. Hier, "Health Information Technology and Electronic Health Records in Neurologic Practice," *Neurologic Clinics*, vol. 28, pp. 411-427, 2010.
- [27] C. Friedman and M. Rigby, "Conceptualising and creating a global learning health system," *International Journal of Medical Informatics*, 2012.
- [28] A. Edwards, I. Hollin, J. Barry, and S. Kachnowski, "Barriers to cross-institutional health information exchange: a literature review," *Journal of healthcare information management*, vol. 24, pp. 22-34, 2010.
- [29] D. Kalra and B. G. Blobel, "Semantic interoperability of EHR systems," *Stud Health Technol Inform*, vol. 127, pp. 231-45, 2007.
- [30] A. L. Rector, "Clinical terminology: Why is it so hard?," *Methods of Information in Medicine*, vol. 38, pp. 239-252, Dec 1999.
- [31] S. T. Rosenbloom, R. A. Miller, K. B. Johnson, P. L. Elkin, and S. H. Brown, "Interface terminologies: facilitating direct entry of clinical data into electronic health record systems," *J Am Med Inform Assoc*, vol. 13, pp. 277-88, May-Jun 2006.
- [32] L. Ohno-Machado, "Biomedical informatics: how we got here and where we are headed," *Journal of the American Medical Informatics Association*, vol. 18, pp. 351-351, Jul 2011.
- [33] H. J. Tange, A. Hasman, P. F. de Vries Robbé, and H. C. Schouten, "Medical narratives in electronic medical records," *International Journal of Medical Informatics*, vol. 46, pp. 7-29, 1997.
- [34] I. Berges, J. Bermudez, and A. Ilarramendi, "Toward semantic interoperability of electronic health records," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 424-31, May 2012.
- [35] G. Cai, "Contextualization of geospatial database semantics for mediating human-GIS dialogues," *Geoinformatica*, vol. 11, pp. 217-237, 2007.
- [36] D. A. Norman, *The Invisible Computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution*. Cambridge, MA: MIT Press, 1998.
- [37] J. J. Cimino, "Desiderata for controlled medical vocabularies in the twenty-first century," *Methods of Information in Medicine*, vol. 37, pp. 394-403, Nov 1998.
- [38] K. H. Hwang, K. I. Chung, M. A. Chung, and D. Choi, "Review of semantically interoperable electronic health records for ubiquitous healthcare," *Health Inform Res*, vol. 16, pp. 1-5, Mar 2010.
- [39] P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt, "Contextualizing ontologies," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 1, pp. 325-343, 2004/10 2004.
- [40] R. M. Turner, "Context-mediated behavior for intelligent agent," *International Journal of Human-Computer Studies, Special issue on Using Context in Applications*, vol. 48, pp. 307--330, 1998.
- [41] M. Berg and E. Goorman, "The contextual nature of medical information," *International Journal of Medical Informatics*, vol. 56, pp. 51-60, 1999.
- [42] V. Kashyap and A. Sheth, "Semantic and schematic similarities between database objects: a context-based approach," *The VLDB Journal*, vol. 5, pp. 276-304, 1996.
- [43] G. Antunes, J. Barateiro, C. Becker, J. Borbinha, and R. Vieira, "Modeling Contextual Concerns in Enterprise Architecture," in *Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2011 15th IEEE International*, 2011, pp. 3-10.
- [44] J. Buck, S. Garde, C. D. Kohl, and P. Knaup-Gregori, "Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach," *International Journal of Medical Informatics*, vol. 78, pp. 521-531, 2009.
- [45] J. Grudin, "Desituating Action: Digital Representation of Context," *Human-Computer Interaction*, vol. 16, pp. 269-286, 2001.
- [46] S. Greenberg, "Context as a Dynamic Construct," *Human-Computer Interaction*, vol. 16, pp. 257-268, 2001.
- [47] C. H. Goh, S. BRESSAN, S. E. MADNICK, and M. D. SIEGAL, "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information," *ACM Transactions on Information Systems*, vol. 17, pp. 270-293, 1999.
- [48] M. Kaenampornpan and E. O'Neill, *Modelling Context: An Activity Theory Approach*, 3295 ed.: Springer, 2004.
- [49] P. Bouquet, F. Giunchiglia, F. v. Harmelen, L. Serafini, and H. Stuckenschmidt, *C-OWL: Contextualizing Ontologies*, 2870 ed., 2003.
- [50] G. Cai and Y. Xue, "Activity-oriented Context-aware Adaptation Assisting Mobile Geo-spatial Activities," in *Proceedings of the ACM IUI 2006: International Conference on Intelligent User Interfaces*, Sydney, Australia, 2006, pp. 354-356.
- [51] A. Analyti, M. Theodorakis, N. Spyrtatos, and P. Constantopoulos, "Contextualization as an independent abstraction mechanism for conceptual modeling," *Information Systems*, vol. 32, pp. 24-60, 2007.
- [52] R. L. Liu and Y. L. Lu, "Context-based online medical terminology navigation," *Expert Systems with Applications*, vol. 37, pp. 1594-1599, Mar 2010.
- [53] E. B. Jakob and B. C. Henrik, "Pervasive Computing Support for Hospitals: An overview of the Activity-Based Computing Project," *Pervasive Computing, IEEE*, vol. 6, pp. 44-51, 2007.
- [54] G. Lakemeyer, "The Situation Calculus: A Case for Modal Logic," *Journal of Logic, Language and Information*, vol. 19, pp. 431-450, 2010.

# A Web-Based System for EEG Data Visualization and Analysis

A. Jonathan Garza, B. Sishir Subedi, C. Yuntian Zhang and D. Hong Lin

Department of Computer Science and Engineering Technology  
University of Houston-Downtown  
Houston, Texas, USA

**Abstract** - While many advances have been made in the understanding of the human brain, it still contains mysteries in its inner workings. The objective of this project is to help discover the answers to more of these mysteries by creating a model that can be applied to Electroencephalographic (EEG) brainwave data to predict what a person is doing or happening to them. The dependent variables of this study are the five major brain waves and the independent variable is the activities performed by the participants. We are creating an environment to capture and analyze EEG brainwave data using various custom developed tools, off the shelf software and hardware components. To help create this environment we are building a website to help facilitate access to and the analysis of collected EEG data. The types of analysis that can be currently performed on the data stored on the web server are wave analysis and statistical analysis. Also a mobile application is being developed to help facilitate the collection of EEG data that is then given to the web server and display the results of data analysis from the web server.

**Keywords:** Electroencephalographic data, Brain State Modeling, Web-Based System, EEG Data Visualization, EEG Data Analysis, computational biology

## 1 Introduction

The human brain is the result of many years of evolution. Its complexities have been studied for years and yet the research done has only begin to scratch the surface. To assist in the aid of that research, a system that can collect, store and processes Electroencephalographic (EEG) data is desirable. This is because EEG signals characterize the result of the neuron activities inside of a human brain. Naturally, they are used to study and understand human brain activities. In particular, EEG signals indicate that neural patterns of meanings in each brain occur in trajectories of discrete steps, whist the amplitude modulation in EEG wave is the mode of expressing meanings [1]. The purpose of the system discussed in this paper is to allow people to easily store, analysis and collaborate on EEG data. The system takes EEG data and exposes it to various analytical techniques so the resultant brain states can be studied and predicted.

The vast implications of using EEG data to analyze brain states include designing brain-computer interfaces (BCI) where users can operate on a machine via brain activities, and using brain state models in healthcare related activities. Scientists now have the ability to measure and register electric potential of the human brain through the use of electroencephalographic technologies. The combination of electroencephalographic data with modeling methods in fields such as data mining and bioinformatics could be used to diagnose disease in advance to increase success of a cure.

It could also be used to prove that subjects in a state of transcendental meditation are in a verifiable and observable state of mind that can be monitored and predicted [2]. Experiments found that cancer patients that practiced meditation experienced higher well-being levels, better cognitive function and lower levels of inflammation than a control group [3].

Therefore, a platform for comprehensive EEG data storage and processing is desirable to promoting applications of using EEG tools in both physiological (e.g., clinical uses, sleep evaluation, fatigue detection, etc.) and psychological (cognitive sciences, BCI, etc.) scopes. Such a platform consists of EEG data collection devices (viz., EEG headset), communication channels (e.g., smart phones), a web server that provides a web interface for users to access stored EEG data and activate data analysis algorithms, an online database for EEG data storage and processing and a forum for users to collaborate with each other while using the system. Figure 1 shows an outline of the proposed system.

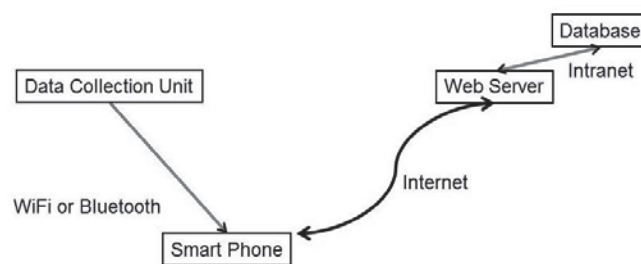


Figure 1. EEG data analysis system architecture

## 2 Related Works

The technology of using the web for visualization purposes has been analyzed before. Nathan Holmberg, Burkhard Wunsch, and Ewan Tempero did a study on Interactive Web-Based Visualization in which they developed a framework to categorized different web-based technologies for 2D and 3D visualization [4]. DHTML which consists of HTML and JavaScript combined, performed well with disadvantages related to limited communication with servers at the time of the writing. Another point made was the popular use of this technology, such as its use by Google Maps. Finally the factor that this technology solely had was the fact that it was not a plugin that users have to install but is the only solution built natively into the web browser. This means users don't need to install any special software to run the visualization software built in DHTML technology.

A study, done by Andrew V. Poliakov, Evan Albright, Kevin P. Hinshaw, et al., found that a major advantage, among others, in a server-client system setup is that the client's hardware does not need to be particular powerful as most of the processing of large data is done on the server side. Servers also tend to be more powerful than personal computers, even at the inexpensive end of servers. Another important factor with regards to server hardware is the popular inclusion of more than one CPU which makes it possible to run parallel data processing methods reducing the overall needed processing time of large amounts of data.

JSON has been shown as a viable way to transmit data from the server to the client software or browser. Many programming languages support JSON messages now and it is the native data representation present in JavaScript which provides convenience when developing website using JavaScript. Web-based system have also shown that it is possible to display multiple records of data together, allowing users to better compare interpersonal difference and similarities between different records as well [5]. This provides a greater aspect of analysis possible than just the displaying of individual records.

Another aspect of web-based system which has shown is that most users found a well-designed system to be easy to use and has a very quick learning curve [6]. This means that non-technical users can easily focus on the analysis of the data and less on the learning on how to use the system.

## 3 EEG Data Collection and Storage

The brain emits electrical signals that are caused by neurons firing in the brain. The patterns and frequencies of these electrical signals can be measured by placing a sensor on the scalp. For example, the EEG sensor by NeuroSky is able to measure the analog electrical signals commonly referred to as brainwaves and process them into digital signals

to make the measurements available for further analysis. Table 1 lists the most commonly recognized frequencies that are generated by different types of brain activity.

### 3.1 EEG Headset

We briefly describe how a simple EEG headset can be built using open source materials. The prototype multi-functional headset we built consists of an EEG sensor, a pulse sensor, a temperature sensor, a microprocessor, and a microprocessor Bluetooth shield.

**Table 1. Brainwave Frequencies.**

Brainwave Type	Frequency Range	Mental States and Conditions
Delta	0.1 Hz to 3 Hz	Deep, dreamless sleep, non-REM sleep, unconscious
Theta	4 Hz to 7 Hz	Intuitive, creative, recall, fantasy, imaginary, dream
Alpha	8 Hz to 12 Hz	Relaxed, but not drowsy, tranquil, conscious
Low Beta	12 Hz to 15 Hz	Formerly SMR, relaxed yet focused, integrated
Midrange Beta	16 Hz to 20 Hz	Thinking, aware of self and surroundings
High Beta	21 Hz to 30 Hz	Alertness, agitation

The assembled headset is shown in Figure 2, where the three sensors are mounted on the tips of the three legs on the forehead supports. The microprocessor and the microprocessor Bluetooth shield are mounted on the back, and the ear lobe is used as an electrical ground base for the EEG sensor.



**Figure 2. Prototype multi-functional**

In order to test and validate that the headset is working properly and that all the sensors are functioning, a test environment had to be constructed. To simulate a real world environment, a mobile smart phone application was developed on the Apple iPhone platform. This platform was chosen for ease of access to development tools and availability of software development kits (SDK) from all the hardware and chipset vendors. Both NeuroSky and Red Bear Labs included sample applications that were then easily transferred to a



custom application using a simple view to display all the sensor values.

### 3.2 Storage

To provide a reliable storage option that can handle the large amount of data from EEG recording sessions, we store the data in a relational database. The database itself is hosted on our web server. The EEG data is stored in its own separate tables apart from other tables necessary for the website. This partly for security reasons but mostly for clarity of which tables are for EEG data and which are not.

The data collected using the NeuroSky headset produces a comma separated values (CSV) file that can be quite large for about 3 minutes of data collection. On average the resulting file is 8 to 9 MB in size which was imported into the database manually in the beginning. To import the EEG data into the database required that the data file first be transferred to the server and then access the database to import the file's contents. We have implemented on the website a method to upload EEG data files seamlessly. After users login, they can choose to upload a new file. Then they will choose which local file they want to upload and the unique name of the data set. After which the file will be uploaded in the background, displaying a progress bar for the user (Figure 3). Once the upload has been completed successfully, the file is stored in an upload folder and then immediately parsed into the database with the name given. Once completed, the EEG data is ready for immediate analysis through the website. If for some reason the upload fails at any point, the user receives an appropriate message. Also on failure of completion, if any incomplete data has been uploaded, it is removed to ensure consistency.

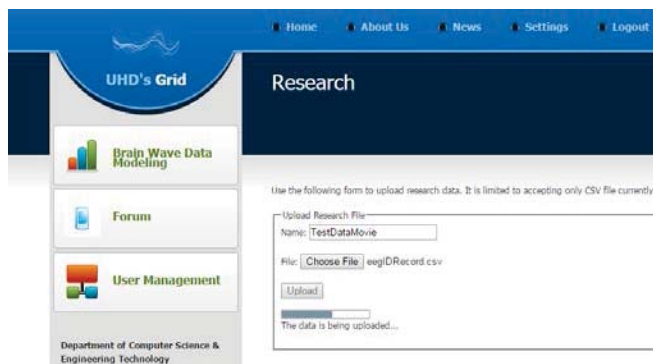


Figure 3. EEG data upload in progress

## 4 Dynamic EEG Data Modeling

To study what EEG data analysis algorithms to be implemented on the web server, we have first begun to do the analysis using Microsoft Excel and the statistical computing programming languages R and Python. Component

frequencies, including five major brain waves- Delta(1-3Hz), Theta(4-7Hz), Alpha Low(8-9Hz), Alpha High(10-12Hz), Beta Low(13-17Hz), Beta High(18-30Hz), Gamma Low(31-40Hz), and Gamma Mid(41-50Hz) were extracted from the raw dataset. These frequencies represent specific brain states including deep meditation and high anxiety.

The data in the headset reports brain wave frequencies as a function of its power spectrum. Fourier Transform analysis was implemented by the application software package to decompose the raw EEG time series into a voltage by frequency spectral graph (power spectrum). This power spectrum values obtained for specific brain waves was investigated for the numerical analysis of Quantitative EEG data.

The power spectrum data was normalized to reduce variability, which might have occurred due to difference in contact distance between the headset and the user, and changes in environmental condition. In addition, automatic scaling feature in the hardware accounts for noises that make the data values large. The normalized method used the sum of all the eight brain waves power spectrum data and divided each data point by the sum to scale it within the range of 0 to 1. The box plot (Figure. 3) graphically depicts the numerical spread of the normalized data for the combined brain waves as well as meditation, movie watching, and reading aloud brain states. The standard deviation model was calculated for each brain state to investigate the variability of data and predict the percentage of values that are present within the one standard deviation from the mean (Figure. 4).

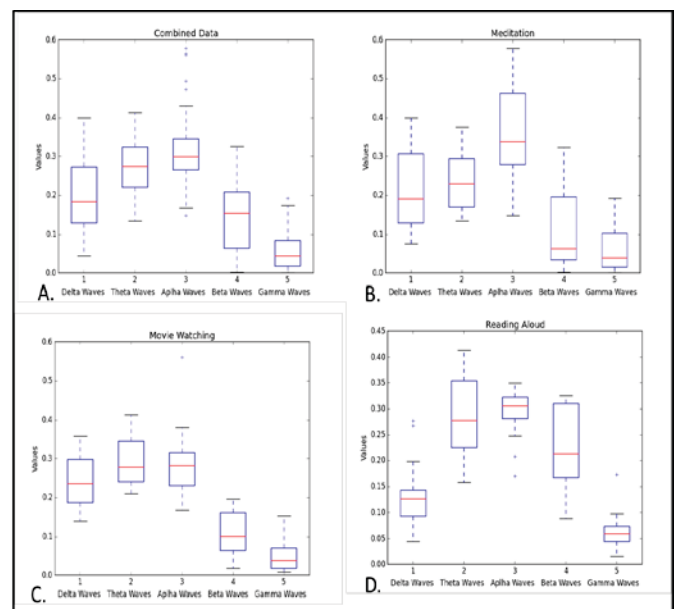
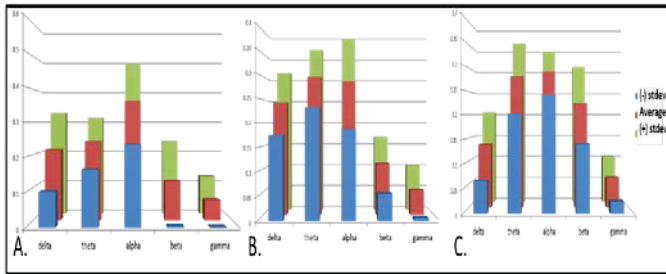


Figure 4. Visualization of normalized data sets using box plot. (A) Combined data (B) Meditation (C) Movie watching (D) Reading aloud

Classification based advanced machine learning algorithm was implemented to further analyze the EEG data from different brain states. The main challenge in this process

was the problem of data separation for each brain wave at different brain states. Generally, the brain waves data from different states tend to cluster together, which becomes tedious for classification algorithms to draw a best fitting separation line.



**Figure 5. Standard deviation model for (A) Meditation, (B) Movie watching, and (C) Reading aloud brain states.**

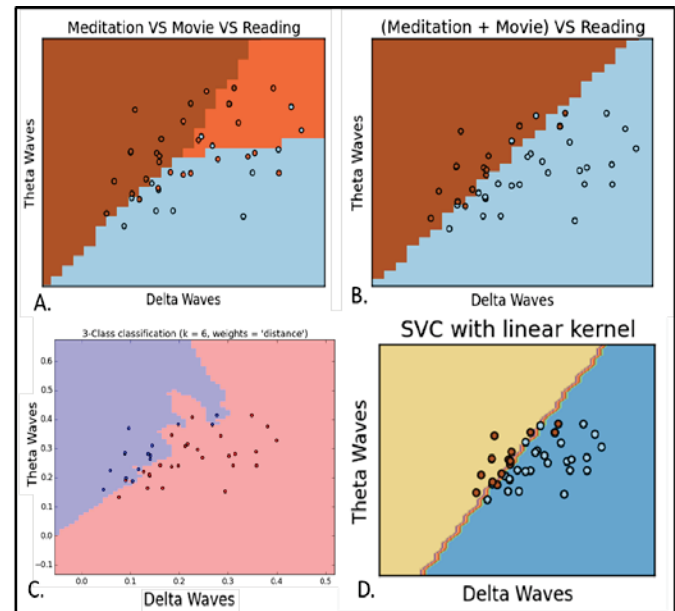
The classification algorithms applied along with the obtained accuracy score is shown in Table. 3. Due to the complexity of data clustering, as an initial step of modeling, two low frequency waves - delta and theta brain waves were chosen as two variables and brain states- mediation, movie watching, and reading aloud were used as three nominal class values discretized as 0, 1, and 2 respectively. The results show that the K-Nearest model contributed the best prediction score with three classes - 78% for mediation, 64% for reading aloud, and 71% for watching a movie. And for two classes - 89% for mediation combined with movie, and 85% for reading brain states.

Interestingly, when two class systems was used by combining meditation and movie watching as class 0, and reading aloud as class 1, the algorithm performed very well with lower error rate (Table. 2). This finding highlights the fact that the majority of volunteers participated for EEG data collection are inexperienced meditator, and the data collected during movie watching and meditation are close to one another than reading aloud. Also, once again with two class system, K-Nearest model performed the best. Since, K-Nearest model depends on highest number of neighboring data point to classify itself to that particular group, the clustering effect of meditation and movie watching should have contributed to the superior performance of the K-Nearest algorithm (Figure 6).

Further, the dynamic fast Fourier transform analysis of EEG data was conducted to reveal the occurrence of dynamic frequency at a steady state along with the time series (Figure. 7). The result shows comparison of FFL graph of three different brain states from two volunteers. In experiences meditator, the graphs show the localized energy of waves whereas in inexperienced meditator volunteer the localization of wave energy is low. This finding underscores the additional layer of complexity for analyzing EEG data.

**TABLE 2. SUMMARY OF CLASSIFICATION BASED PREDICTION SCORES**

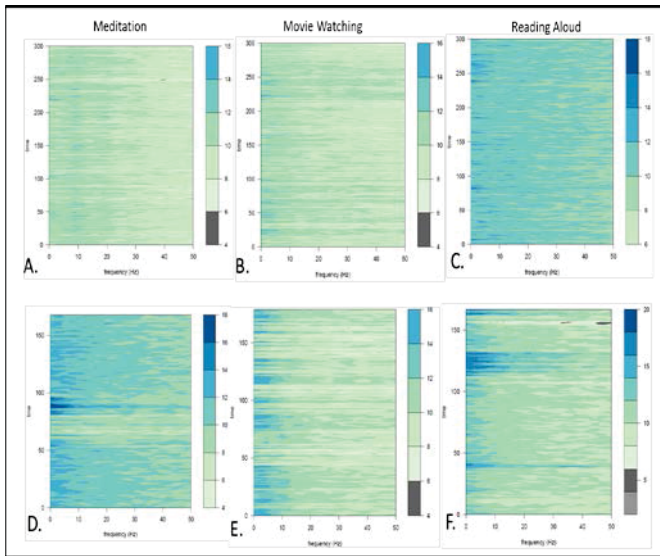
Model	Brain State	Standard Deviation Model	K Nearest Neighbor	Support Vector Machine	Naive Bayes	Logistic Regression	Multiclass Algorithm
Three Class	0. Meditation	64.29	78.57	78.57	50.00	78.57	50.00
	1. Movie Watching	75.71	71.43	57.14	71.43	21.43	35.71
	2. Reading Aloud	65.71	64.29	14.29	78.57	57.14	85.71
Two Class	0. Meditation and Movie Watching	-	89.29	82.14	-	85.71	-
	1. Reading Aloud	-	85.71	64.29	-	78.57	-



**Figure 6. Data analysis of brain waves. Standard deviation model for (A) Logistic Regression model three class (color: blue-meditation, orange-movie, and brown-reading), (B) Logistic Regression model two class (color: blue-meditation plus movie, and brown-reading), (C) K-Nearest model two class (color: pink-meditation plus movie, and blue-reading), and (D) Support Vector Machine with Linear Kernel two class (color: blue-meditation plus movie, and yellow-reading).**

## 5 Web Visualization of EEG Data

We want to take a new approach through affective computing, which employs EEG signals recorded when users perform some brain activities and apply analytical algorithms to captured EEG data to detect the brain state. EEG signals can be measured at any moment and are not dependent on feelings, emotions, or human behaviour. We are investigating an automatic EEG-based recognition system that can record the EEG signals from users and measure their brain states. The EEG data are filtered to get separate frequency bands which are then analysed then displayed via a web server and web user interface.



**Figure 7. Dynamic Fourier analysis for two volunteers (A-B-C: Inexperienced Meditator, D-E-F: Experienced Meditator) during meditation, movie watching, and reading aloud brain states.**

**5.1 Web Interface Visualization**

The web server provides a user interface that allows users to view EEG data in the database and perform data analysis. Figure 8 and Figure 9 show the web interface of the data which is rendered in wave form mode and statistical mode, respectively.

The wave form rendering seen in Figure 8, is generated dynamically within the user’s browser. This allows the graph to be zoomed in (Figure 10) or moved around to allowed focus on a specific section of the graph.

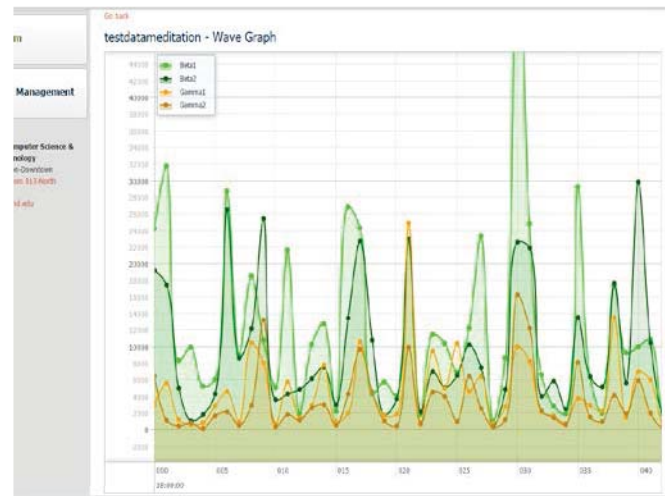
**5.2 Mobile Application**

We have also developed an iPhone app. iPhone users can use the app that will connect to the database to view data. Figure 5 shows the two functions, viz., “collect data” and “view data”, that a user can choose on the iPhone app. The user can display data in text mode by viewing individual data frames (Figure 11(a)), or display the wave form of recorded data in certain time period (Figure 11(b)).

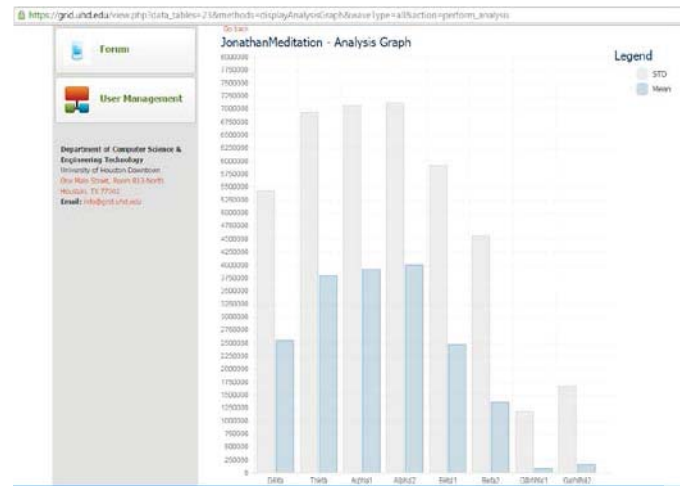
**5.3 Web Collaboration**

We have developed a forum to provide a way for researchers to communicate with each other through the website. The forum has been built using phpBB because of it is open source software and also because of its features. The forum has been divided into categories to better focus discussions within a category (Figure 12). The categories are currently Website, Modeling, and Off-Topic. Users can then post their message or reply within the appropriate category for other members to see. Users can also include in their post

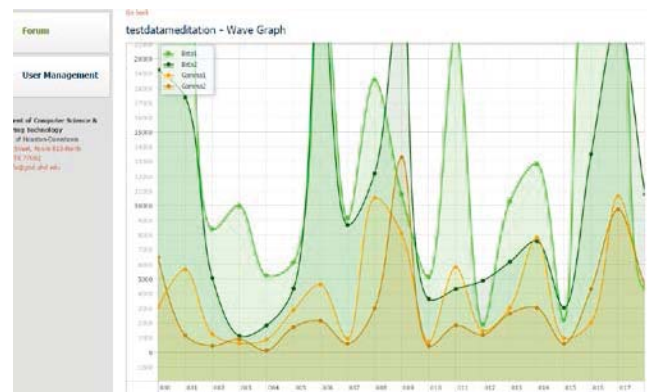
pictures or files that might help further there discussion. For example, they might include in their post an image of a wave graph from an analysis of their data from our website to discuss with other members.



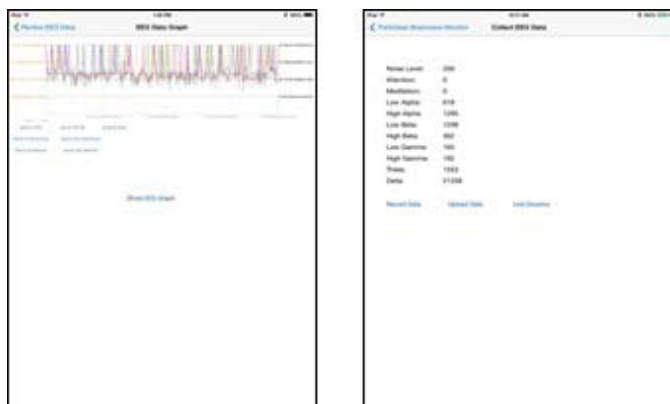
**Figure 8. EEG Wave Forms Rendering in Web Interface.**



**Figure 9. EEG Histogram Rendering in Web Interface.**



**Figure 10. Zoomed-in version of the EEG wave rendering**



(a) Mobile app wave form graph visualization (b) Text mode display

Figure 11. Mobile iPhone app

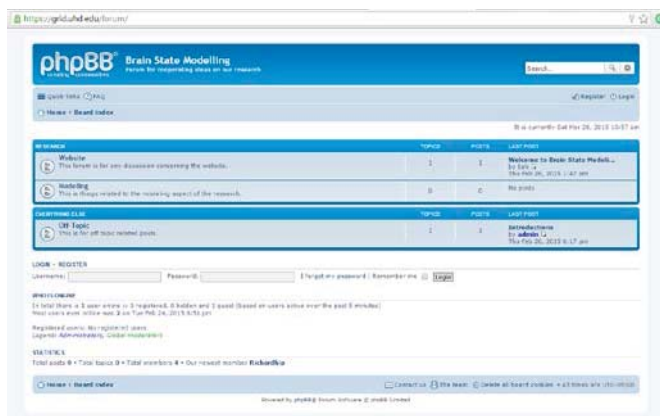


Figure 12. Website's Forum

## 6 Conclusions

The EEG headset, mobile application, web server, and web interface proved to be a good starting foundation for a proof of concept setup and to be able to visually see the EEG data capture and display from start to finish. It showed that it would be possible to capture EEG data from anywhere, while on the move, and be able to immediately see some of the results. Although much more refinement is required in increasing the performance of the data capture and reducing the electrical noise interference during the EEG capture, the system as built provides a good foundation for future improvements.

## 6 Future Works

The system that we have developed currently shows great promise for future implementation of more analysis algorithms. We are working on incorporating more of those analysis algorithms after developing and testing them first locally before inclusion on the website. We will also develop

more visualizations methods for the website as necessary for different algorithms so as to represent results in the best visual appropriate. We also will further develop the mobile application to be more tightly integrated with the website. We also aim to address the complexity of classification of brain waves data by modeling the major brain waves independently with clinically significant brain regions combined with the time-series analysis. This will achieve an efficient and predictable brain wave modeling system which has potential application in hospitality and clinical industry for self-controlled deep brain relaxation and early diagnosis of various brain abnormalities respectively.

## 7 Acknowledgements

Gregor Schreiber and Minghao Yang contributed to the work presented in this paper. This work is partially supported by NSF CAHSI grant, NSF UGI-CSTEM grant, and NSF Summer REU grant.

## 8 References

- [1] W.J. Freeman, "A neurobiological interpretation of semiotics: meaning, representation, and information," *Information Sciences*, 124(2000), 93-102, 2000.
- [2] R. Davidson, J. Kabat-Zinn, J. Schumacher, et al., "Alterations in Brain and Immune Function Produced by Mindfulness Meditation," *Psychosomatic Medicine*, 65(4):564-570, 2003.
- [3] B. Oh, P. Butow, B. Mullan, et al., "Impact of medical Qigong on quality of life, fatigue, mood and inflammation in cancer patients: a randomized controlled trial," *Annals of Oncology*, (3):608-614, 2009.
- [4] Nathan Holmberg, Burkhard Wunsche, and Ewan Tempero. "A framework for interactive web-based visualization"; *AUIC '06 Proceedings of the 7th Australasian User interface conference* (Australian Computer Society, Inc.), 50, 137-144, Jan. 2006.
- [5] Andrew V. Poliakov , Evan Albright , Kevin P. Hinshaw , David P. Corina , George Ojemann , Richard F. Martin , James F. Brinkley. "Server-based Approach to Web Visualization of Integrated Three-dimensional Brain Imaging Data"; *Journal of the American Medical Informatics Association*, 12, 2, 140-151, Mar 2005, DOI: 10.1197/jamia.M1671
- [6] Lourenço, A., Plácido da Silva, H., Carreiras, C., Priscila Alves, A., & L. N. Fred, A. "A web-based platform for biosignal visualization and annotation"; *Multimedia Tools and Applications* (Springer US), 70, 1, 433-460, May 2014.

# Using Temporal Logic to Verify Blood Supply Chain Safety

Noha Hazzazi, Bo Yu, Duminda Wijesekera, and Paulo Costa

{nhazzazi|byu3|dwijesek|pcosta}@gmu.edu.

CARE, Volgenau School of Engineering, George Mason University, Fairfax, VA, 22030

**Abstract**—*The need for transfusion blood increases each year. The Food and Drug Administration (FDA), the American Association of Blood Banks (AABB) and standardization bodies that hold jurisdiction in other countries continuously update blood transfusion safety mandates. Verifying blood bank processes for safety takes labor and time. We automate this verification process by modeling the work flows of the blood processing supply chain, extracting FDA and AABB requirements as Temporal Logic formulas and verifying that the workflows comply with the mandates. We also show how this process can seamlessly integrate into an Electronic Medical Record System.*

**Keywords:** Blood bank, Verification, Temporal Logic

## 1. Introduction

Blood transfusion is a usual procedure for patients suffering major surgeries, injuries or illnesses such as hemolytic anemia. Each year, almost five million Americans need a blood transfusion [1]. Transfusions are administered in 137 countries serving a population of 3.1 billion worldwide [2]. The healthcare industry faces increasing challenges for ensuring safe, quality blood transfusions, including:

- Risk of transmission of infection through unsafe blood or blood products
- Technical and clerical errors in the processing and testing of blood
- Errors in the administration of blood or blood products.

The World Health Organization (WHO), in its Strategy for Blood Safety and Availability for Improving Patient Health and Saving Lives, says that cumulatively 58% of transfusion fatalities are caused by transfusion errors and related issues [3]. Transfusions are the tip of a long, well-regulated supply chain that consists of collecting blood from donors; testing and decomposing blood into components such as red blood cells, platelets etc.; preserving them in regulated conditions (e.g. adding preservatives and refrigeration); and conducting the transfusion based on physicians' orders. The complexity of this chain results in many potential causes of unsafe blood transfusions.

In past decades, many researchers and organizations transferred efforts and resources from creating standards to adopting new IT technologies that would improve blood transfusion safety. For example, the WHO as well as the AABB among others defined blood safety standards [2]. In

the US, blood banks must follow FDA blood safety standards [4]. To make the transfusion chain safer, many health organizations adopted blood bank systems or other computerized systems, such as the TANGO Automated Blood Bank System, SIBAS [5]. Adverse effects of transfusions are mostly examined and appropriate corrective actions are taken routinely.

In most countries, blood product safety relies on manual documentation that are not linked together, such as paper-based or computerized forms scattered throughout the blood supply chain. Blood bank systems have moved from manual documentation to complex systems with improved functionality, such as inventory control and policy management [6], [7]. However, current systems assist in collecting the necessary documentation in electronic formats but do not provide electronically verifiable safety assurance in each step and for the supply chain as a whole as specified by each locality's safety standards. We are unaware of any system that uses safety-verified workflows in administering blood products. Providing a method to do so is the objective of this paper.

Observed errors committed by blood bank staff, especially by limited staff during night shifts who nonetheless handle normal workload [8], compels electronically enforceable blood safety verification. We propose to introduce a safety verification method that can be built into existing Electronic Medical Record and blood bank systems using established engineering practices. Our approach includes formally modeling the supply chain of blood as a workflow, specifying the safety regulations in a logical format, and verifying that the appropriate components of the entire supply chain satisfy all applicable safety regulations. This paper shows an early prototype of our proposal.

This paper is divided as follows. Section 2 describes our process to verify the safety of the supply chain of blood, which starts with modeling the supply chain as a workflow system, which states are formed by the input/output of participants - or processes working on behalf of them - resulting into succeeding states (Figure 2). Then, in Section 3 FDA and AABB safety regulations are translated to Temporal Logic statements that have to hold in particular states of the blood processing supply chain. We use a model checker to verify these statements and whether the states on the supply chain satisfy the appropriate Temporal Logic Formulas. Section 4 shows how this entire process is transparent to an EMR system using an open source tool. Finally, Section 5 describes related work and Section 6

presents our conclusions.

## 2. The Blood Safety Workflow

Figure 1 shows our waterfall model of collecting information, remodeling, validating, verifying, and implementing the blood safety workflow-enforced blood bank system.

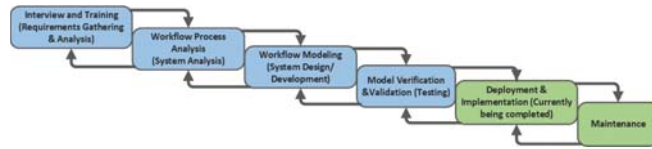


Fig. 1  
RESEARCH METHOD [9], [4], [10].

This section covers our model of the high level view of the FDA regulations and AABB standard. We interviewed blood bank professionals and created the complete workflow shown in Figure 2. The diagram shows all composite processes in filled color, with their associated decomposition contouring the same color. The main processes are (1) Registration (2) Physical Exam (3) Collection (draw blood) (4) Post Donation, and (5) Transfusion. All processes, from beginning of donation until transfusion, abide to the safety requirements set forth by the FDA and AABB. To the best of our knowledge, there is no published blood bank workflow that has been verified against the FDA and AABB requirements.

The first main process is Registration (four dark cyan color boxes at the top left of Figure 2), which captures donor demographics prior to Physical Exam. Its specification requires the donor to be identified, registered, and provided with educational material, all stemming from a safety measure mandate of tracking all processes from registration until patient follow-ups [11]. For instance, the detailed procedure in the "Identify & Verify Donor" step (second box from top) involves identifying the donor by matching its ID photo with the person, and verifying its expiration date and validity.

The second step is the Physical Exam (Aqua-colored boxes in Figure 2), which determines the donor's suitability for donating blood. This composite process depends on the type of donation; as there are separate regulatory requirements for whole blood or apheresis donation. To illustrate the process we will focus on high level whole blood suitability.

Whole blood donors are required to check their hemoglobin level, blood pressure, temperature, arm diseases, and donation interval. These measures are specified by the FDA to ensure donor and blood safety. For instance, in CFR 640.3 the FDA specifies (1) normal temperature and (2) systolic and diastolic blood pressures are within normal limits. However, no quantifiable definition of normal is provided. For this reason, many blood banks apply other standards (e.g. AABB) in support to the FDA's specification. The AABB, for example, defines thresholds of  $\leq 37.5^{\circ}\text{C}$  for temperature,

$\leq 180\text{mmHg}$  for Systolic pressure, and  $\leq 100\text{ mm Hg}$  [9] for Diastolic pressure. Regarding the other requirements for whole blood, FDA specifies that I donor must (a) have its hemoglobin  $\geq 12.5$  grams, (b) be free from skin diseases or arm scars, and (c) from diseases transmissible by blood. In our model, we defined the properties of the physical exam/suitability process based on these requirements.

Once eligibility is established, the donor is ready for the Collection (Cadet Blue-colored boxes in Figure 2). Collections also has specific requirements for regular donations (whole blood) and apheresis, which covers aspects such as unit identification, donor arm preparation, temperature control. Whole blood and apheresis require each unit to be linked to a donor via his/her donor ID. In CFR 640.4, FDA states that the skin of the donor at the site of phlebotomy shall be prepared thoroughly and carefully by a method that gives maximum assurance of a sterile container of blood and a proper donor arm preparation to prevent blood contamination. Temperature control after the blood unit is collected is another requirement, with the FDA defining that collected blood must be placed in storage at a temperature between 1 and  $6^{\circ}\text{C}$  immediately after whole blood collections. The FDA CFRs 640.4, 640.14, 640.22, 640.32 state information about whole blood collection but are not clear on either apheresis collection or on the volume of blood collected. For this reason, we resorted to the AABB specification for the volume of blood required on whole blood and apheresis collections (5.6.4 and 5.5.3.5, respectively, in [12]).

After collection, the next step is post donation consisting of blood processing and donor monitoring (Fuchsia-colored boxes in Figure 2). As stated by blood bank professionals, the donor is monitored to ensure he/she did not develop any adverse reactions to ensure his/her safety. It is mandated to collect extra samples of the donated blood unit at the time of collection for testing (modeled in Figure 2). CFR 640.5 states that all laboratory tests shall be made on a specimen of blood taken from the donor at collection time.

CFR 640.5 also specifies the main tests required to process the blood unit, as well as its acceptable results. This includes tests such as serological test for syphilis, blood grouping and Rh tests. Our model in Figure 2 computes the FDA mandated requirement for the method used for blood grouping. This is stated in CFR640.5 part (b) as *at least two blood group tests shall be made and the unit shall not be issued until grouping tests by different methods or with different lots of antisera are in agreement*. Regarding Rh grouping, most blood banks currently carry out both the blood grouping and Rh in parallel, which is also captured in our model.

Further tests must be done for sterility, blood unit inspection, and test for communicable disease based on CFR 640.5 (Sections d, e, f). Blood used for transfusion should not be tested for sterility, but we did not model this aspect since our focus this time was on usable blood. All donated blood shall be visually inspected and the FDA states that it

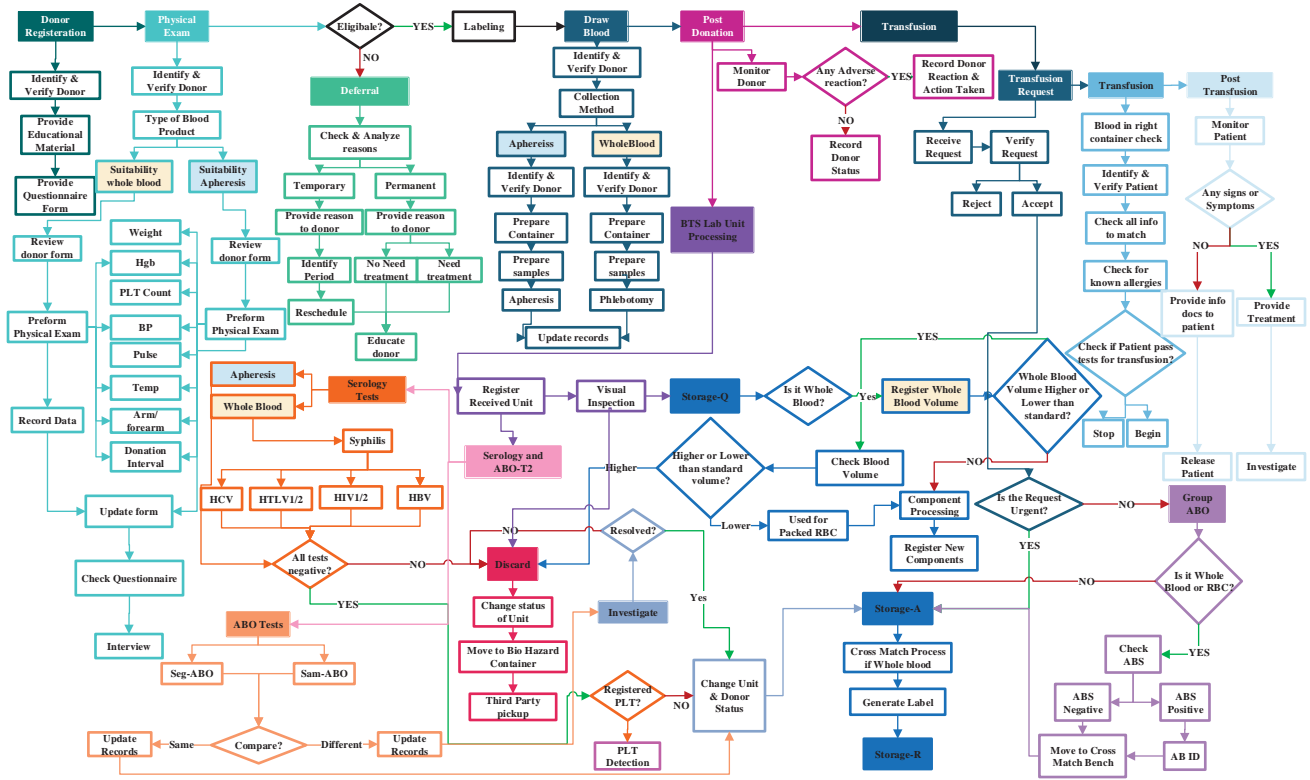


Fig. 2

HIGH LEVEL VIEW OF THE FDA REGULATIONS FOR THE BLOOD SUPPLY CHAIN [9], [4], [10].

should be inspected at storage and prior to using. The list of communicable diseases required to check is covered in CFR 610.40 and included in our model.

Once blood is collected, tested and stored, the next step is Transfusion (Midnight Blue-colored boxes in Figure 2), in which an authorized health professional requests blood transfusion, transfuses the compatible blood and monitors the patient. The AABB have specified the details required to request for blood transfusion (cf. parts 5.11.1.1 and 5.28.2 [12]) and for checking blood group ABO/Rh (cf. parts 5.12 and 5.12.1 [12]), which are both captured as processes shown in Figure 2.

### 3. Temporal Logic Translation

In this section, we show how we translate FDA and AABB mandated safety requirements. The verification process consists of defining the predicates, states, and properties in order specify and verify safety requirements. We start this step by translating the FDA and AABB safety requirements into Linear temporal Logic (LTL) statements.

### 3.1 Syntax

In order to specify LTL syntax, let  $VAR = \{\vec{x}_i; i \geq 0\}$  be a set of variables,  $CONST = \{\vec{c}_i; i \geq 0\}$  be a set of constants and  $\Phi = \{p_i; i \geq 1\}$  be a set of atomic predicate symbols.

We say that  $p_i(\vec{x}_{i_j}), p_i(\vec{x}_{i_j}) \wedge p_k(\vec{x}_{k_j}), p_i(\vec{x}_{i_j}) \vee p_k(\vec{x}_{k_j}), \neg p_i(\vec{x}_{i_j}), \exists \vec{x}_{i_j} p_i(\vec{x}_{i_j}), \forall \vec{x}_{i_j} p_i(\vec{x}_{i_j}), p_i(\vec{x}_{i_j}) \rightarrow p_k(\vec{x}_{k_j})$  and  $\diamond p_i(\vec{x}_{i_j}), \square p_i(\vec{x}_{i_j}), \mathcal{X} p_i(\vec{x}_{i_j})$  (sometimes this next-time operator  $\mathcal{X}$  is written as  $\bigcirc p_i(\vec{x}_{i_j})$ ) are predicates. Following standard convention, a fully instantiated predicate is one in which all variables are replaced by constants where we write  $p_i(\vec{c}_{i_k}) \vec{x}_{i_j}$  to indicate that the variables  $\vec{x}_{i_j}$  in  $p_i(\vec{x}_{i_j})$  have been replaced with constants  $\vec{c}_{i_k}$ .

### 3.2 Semantics

We now summarize the commonly used semantics of temporal logic. Let  $S = \{s_i; i \geq 0\}$  be a collection of states (sometimes referred to as worlds) and an accessibility relation among states as  $R \subseteq S \times S$ . We assume that there is a mapping (referred to as an assignment of the fully instantiated instances of the predicate symbols), say  $Inst = \{inst_k k \geq 0\}$  with the mapping  $AtMap = M :$

$Inst \mapsto \wp(S)$ . Then we define the satisfaction relations for the predicates in the states as follows:

- $s_i \models inst_k$  if  $s_i \in AtMap(in_k)$ .
- $s_i \models inst_k \wedge inst_j$  if  $s_i \models inst_k$  and  $s_i \models inst_j$ .
- $s_i \models \neg inst_k$  if  $s_i \notin inst_k$ .
- $s_i \models inst_k \vee inst_j$  if  $s_i \models inst_k$  or  $s_i \models inst_j$ .
- $s_i \models \forall x p_k$  if  $s_i \models inst_k$  for every instance  $inst_k$  of  $p_k$  and the only free variable of  $p_k$  is  $x$ .
- $s_i \models \exists x p_k$  if  $s_i \models inst_k$  for some instance  $inst_k$  of  $p_k$  and the only free variable of  $p_k$  is  $x$ .
- $s_i \models \diamond inst_k$  if  $s'_i \models inst_k$  for some  $s'_i \in R^*(s_i)$ , where  $R^*$  is the reflexive transitive closure of  $R$ .
- $s_i \models \square inst_k$  if  $s'_i \models inst_k$  for every  $s'_i \in R^*(s_i)$ , where  $R^*$  is as stated above.
- $s_i \models \mathcal{X} inst_k$  if  $s'_i \models inst_k$  for some  $s'_i \in R^*(s_i)$ .

The sample version described in the paper uses only seven states  $S = \{s_1, s_2 \dots s_7\}$ . We also use 66 predicates, i.e.  $P = \{t_1, t_2 \dots t_{66}\}$ , seven constants. Our model checker uses  $X$  for  $\circ$ . Sample safety requirements shown in this paper currently do not use the connective  $\square$ . We show verification related to the registration, donor suitability and the main workflow.

### 3.3 Mapping Safety Requirements as Assertions in States and State Transitions

We describe the state transitions and safety properties that must be satisfied by state transitions as per FDA and AABB specifications. Given the complexity of the process, we modeled the workflow as consisting of sub-workflows and sub-sub workflows. These result in having hierarchical states and safety assertions associated with transitions between them, which we write as temporal logic formulas and verify using a model checker. For the purpose of this paper, we described sample state transitions, how we decomposed FDA and AABB requirements into LTL assertions about hierarchical states, and how we verified them. Table 1 provides a summary of our assertions.

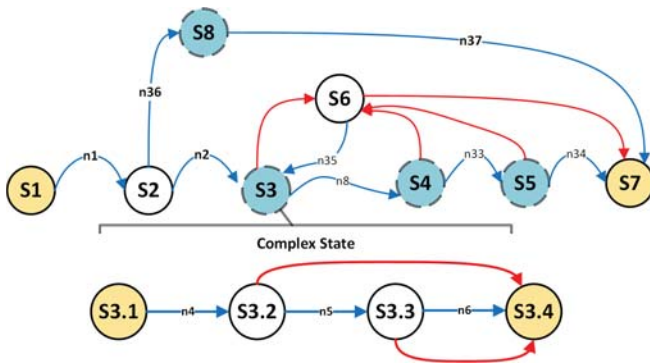


Fig. 3  
ATOMIC STATES.

As shown in Figure 3, state begin at  $S_1$  and safety properties are shown in Table 1.

- $S_1$ : The user starts the workflow by entering *start* into the system. On trigger  $t_1$  (the user entering *start*) a transition  $n_1$  takes the system from  $S_1$  to  $S_2$  as shown in Figure 3.
- $S_2$ : The user gets donor's demographics [DonorName, DonorAge, NationalSA, GovID, Address, Occupation, Homephone, Mobile, Email, MartialStatus, Gender, Volunteer Autologous, FirstTimeDonation, Donation-TypeRegular]. Based on the entered demographics the system routes the user between *USregistration* or *SAregistration*. (In this paper we will cover only *SAregistration* and *SAsuitability* that corresponds to the registration process in a Saudi hospital, modeled by states  $S_3$  and  $S_4$ ). On trigger of  $t_4$  (it passes the nationality such as Canada, US, New Zealand, Saudi Arabia, China), the state in transition  $n_2$  from  $S_2$  to  $S_3$ . Conversely, if other nationality state transitions from  $S_2$  to  $S_8$  (shown in transition  $n_{36}$ ).
- $S_3$ : In this composite state, the system checks the donor demographic information. It decomposes the sub workflow (into  $S_{3.1}$ - $S_{3.4}$ ) and outputs a Boolean flag *VerifiedID*. It is modeled as transition  $n_{32}$  that on trigger  $t_2$  (user identify and verify the donor ID) transition from  $S_3$  to  $S_4$ . Conversely, if the donor ID is not verified, transition  $n_{35}$  takes  $S_3$  to  $S_6$ , as shown in Figure 3.
- $S_4$ : In this composite state, the system checks if the donor is suitable for the donation and output a Boolean flag *DonorSuitable*. This state decomposes the tasks into sub workflows (going from states  $S_{4.1}$  to  $S_{4.8}$ ) and outputs a Boolean flag *DonorSuitable* indicating the donor passing the physical exam. It is modeled as transition  $n_{33}$  that on trigger  $t_3$  (user identify the donor suitability) transition from state  $S_4$  to  $S_5$ .

Now we briefly describe the decomposition of a few complex states. Compound state *State 3* in Figure 3 is decomposed into eight sub-states ( $S_{3.1}$  to  $S_{3.8}$ ) but we describe only three states in this paper. The second row in Table 1 shows the safety property of validating the ID of a new donor should hold in state  $S_{3.1}$ . In the beginning state of sub-workflow the system imports the donor demographics [GovID] from the main workflow. When the donor is successfully verified, transition  $n_4$  takes the system from  $S_{3.1}$  to  $S_{3.2}$ , as shown in Figure 3.

We show the decomposition of compound state *state 4* in Figure 4 where the safety property associated with the third row of Table 1 holds in state  $S_{4.1}$ . Because  $S_{4.1}$  is the beginning state of sub-workflow where the system imports the donor demographics [GovID, DonorAge, Autologous and FirstTimeDonation] and starts as state  $S_4$ . The decomposition takes the system through 14 states, numbered  $S_{4.4.1}$  through  $S_{4.4.14}$  of which we describe 5. It is modeled as



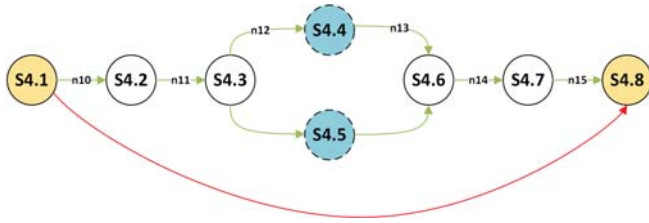


Fig. 4  
COMPOUND STATES IN  $S_4$

transition  $n_{10}$  that is triggered by action *start* and goes from state  $S_{4.1}$  to  $S_{4.2}$  in Figure 4. We only describe two states of this sub-sub-workflow and associated safety properties for brevity.

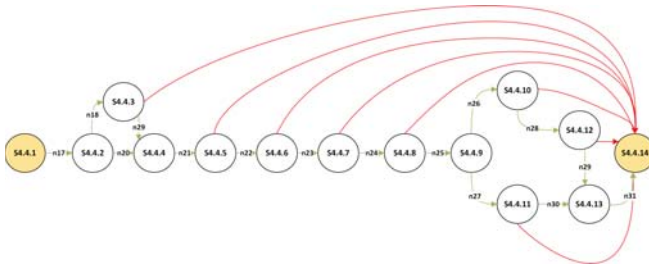


Fig. 5  
COMPOUND STATES IN  $S_{4.4}$

We show the decomposition of compound state *state 4.4* in Figure 5 with the fourth property holding at  $S_{4.4.1}$  explained in Table 1:

- $S_{4.4.1}$ : The start state of the sub-workflow, where the user clicks the start button to begin the sub-workflow. This transition  $n_{17}$  takes the system from state to  $S_{4.4.1}$  to state  $S_{4.4.2}$  triggered by the action *start*.
- $S_{4.4.2}$ : In this state, the system imports [FirstTimeDonation] and checks if the donor had previous donations, modeled as the transition  $n_{18}$  that takes the system from state  $S_{4.4.3}$  to state  $S_{4.4.3}$  triggered by the action  $t_{57}$ .
- $S_{4.4.3}$ : The user enters [DonationIntervalDays and DonorSuitable] to route the donor. It is modeled as transition  $n_{19}$  takes the system from state  $S_{4.4.3}$  to state  $S_{4.4.3}$  triggered by actions  $t_{63}$ .
- $S_{4.4.4}$ : The user enters the donor [Temp, Syspressure, Diapressure, Armclear, Hemoglobin], as the blood bank technician performs the physical exam. This is modeled as transition  $n_{20}$  that takes the system from state  $S_{4.4.4}$  to state  $S_{4.4.5}$  automatically.
- $S_{4.4.5}$ : The system checks if the entered temperature is in the normal range. In order to do so, the system imports [Temp] into the process and ask the user to enter [Donorsuitable]. This is modeled by the transition

$n_{21}$  that takes the system from state  $S_{4.4.4}$  to state  $S_{4.4.5}$  triggered by action  $t_{66}$ .

## 4. Implementation

This section describes our implementation as shown in Figure 6, which extends previous work [13], [14] by adding a workflow specification verification component. The extended system includes three components. (1) Electronic Medical System, (2) Workflow Management System (WFMS), and (3) Workflow Specification Verification System. For (1) we adopt an open-source EMR system, OpenMRS, as our EMR System component. We create the blood bank user interfaces to be used by blood banks. All patient data is stored in OpenMRS databases. For (2) we have Workflow Editor and Workflow Engine. Workflow editor models the comprehensive, completed blood bank workflow that we created for collecting and administering blood products. Workflow Engine enforces such workflow, which takes blood bank professionals through safety verified processes when collecting and administering blood products. We also use an open-source WFMS—namely, YAWL—as our WFMS, which is implemented as a loadable module in OpenMRS[15]. For (3), which is highlighted in yellow in Figure 6), the two main components are Workflow Specification Translator and a Model Checker. The Workflow Specification Translator creates Divine (DVE) syntax of the workflow specifications. The Model Checker verifies the model for safety requirements specified as LTL properties. As stated previously, we encode FDA and AABB blood safety requirements to ensure that our workflow model is compliant. Finally, WFMS databases store the verified blood bank workflow specification for collecting and administering blood products.

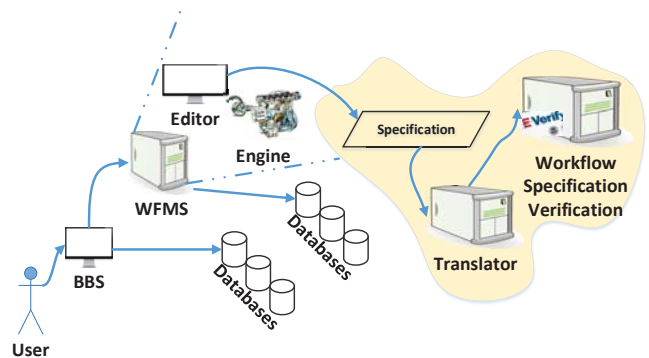


Fig. 6  
IMPLEMENTATION ARCHITECTURE

Table 1  
STATES, PROPERTY AND PREDICATES SUMMARY.

State	Property in English	Property in Predicates	Property in LTL Syntax
S1	Checks if the nationality is set properly, the ID is verified.	Next NationalitySA $\neq \perp$ and two states after VerifiedID $\neq \perp$ or VerifiedID $\neq \top$ and three states after TryAnotherID $\neq \perp$ and five states after DonorSuitable $\neq \perp$	<ul style="list-style-type: none"> <li>In the Model Checker #property <math>X(t_4 \ \&amp;\&amp; \ X(t_2 + (!t_2 \ \&amp;\&amp; \ Xt_5)) \ \&amp;\&amp; \ XXt_3 \ \&amp;\&amp; \ XXXt_3)</math></li> <li>In LTL <math>\bigcirc(t_4 \ \&amp;\&amp; \ \bigcirc(t_2 + (!t_2 \ \&amp;\&amp; \ \bigcirc t_5)) \ \&amp;\&amp; \ \bigcirc\bigcirc t_3 \ \&amp;\&amp; \ \bigcirc\bigcirc\bigcirc t_3)</math></li> </ul>
S3.1	Checks if the donor is new and if so validate the d ID, by checking the ID Expiry Date and Valid Photo.	Next NewDonor $\neq \perp$ and two states after it checks for ValidID $\neq \perp$ and ValidIDExpiryDate $\neq \perp$ and ValidPhotoID $\neq \perp$	<ul style="list-style-type: none"> <li>In the Model Checker #property <math>X(t_{13} \ \&amp;\&amp; \ X(t_{14} \ \&amp;\&amp; \ t_{15} \ \&amp;\&amp; \ t_{16}))</math></li> <li>In LTL <math>\bigcirc(t_{13} \ \&amp;\&amp; \ \bigcirc(t_{14} \ \&amp;\&amp; \ t_{15} \ \&amp;\&amp; \ t_{16}))</math></li> </ul>
S4.1	Validate the donor identity by checking if the donor is holding a valid ID, Valid ID expiry date and valid photo ID.	Next ValidID $\neq \perp$ and ValidIDExpiryDate $\neq \perp$ and ValidPhotoID $\neq \perp$ and next next DonationTypeRegular $\neq \perp$	<ul style="list-style-type: none"> <li>In the Model Checker #property <math>X((t_{53} \ \&amp;\&amp; \ t_{54} \ \&amp;\&amp; \ t_{55}) \ \&amp;\&amp; \ Xt_{24})</math></li> <li>In LTL <math>\bigcirc(t_{53} \ \&amp;\&amp; \ t_{54} \ \&amp;\&amp; \ t_{55}) \ \&amp;\&amp; \ \bigcirc t_{24})</math></li> </ul>
S4.4.1	Check if this is the donor's first donation. If donated blood before and the time interval between the previous donation and now is $\geq 57$ days the donor can donate. Then the the system ensures that the temperature is $\leq 37.5^\circ C$ . In five successive states the system will ensure that the donors systolic pressure is $\leq 180$ and, the diastolic pressure $\leq 100$ and the donor has a disease free arm.	Next FirstTimeDonation $\neq \perp$ and next next DonationIntervalDays $\geq 57$ and next next next next Temp $\geq 37.5$ and next next next next next SysPressure $\leq 180$ and next DiaPressure $\leq 100$ and next ArmClear $\neq \perp$ and next DonorAge $\geq$ and next Autologous $\neq \perp$ or next Autologous $\neq \top$ and next DonorSuitable $\neq \perp$	<ul style="list-style-type: none"> <li>In the Model Checker #property <math>X(t_{57} \ \&amp;\&amp; \ X(Xt_{66} + (t_{63} \ \&amp;\&amp; \ XXt_{66})) \ \&amp;\&amp; \ (Xt_{65} \ \&amp;\&amp; \ Xt_{62} \ \&amp;\&amp; \ Xt_{61} \ \&amp;\&amp; \ X((t_{58} \ \&amp;\&amp; \ Xt_{64} \ \&amp;\&amp; \ Xt_{59} \ \&amp;\&amp; \ Xt_{60})) + (X!t_{58} \ \&amp;\&amp; \ Xt_{64} \ \&amp;\&amp; \ Xt_{60})))</math></li> <li>In LTL <math>\bigcirc(t_{57} \ \&amp;\&amp; \ \bigcirc(\bigcirc t_{66} + (t_{63} \ \&amp;\&amp; \ \bigcirc\bigcirc t_{66})) \ \&amp;\&amp; \ (\bigcirc t_{65} \ \&amp;\&amp; \ \bigcirc t_{62} \ \&amp;\&amp; \ Xt_{61} \ \&amp;\&amp; \ \bigcirc((t_{58} \ \&amp;\&amp; \ \bigcirc t_{64} \ \&amp;\&amp; \ \bigcirc t_{59} \ \&amp;\&amp; \ \bigcirc t_{60})) + (\bigcirc \neg t_{58} \ \&amp;\&amp; \ \bigcirc t_{64} \ \&amp;\&amp; \ \bigcirc t_{60})))</math></li> </ul>

We describe the verification of our model against safe blood requirements, highlighted in yellow in Figure 6.

## 5. Related Work

Systems such as SCC Soft Computer, MAK-system and many others [16], [17] do not formally model and verify blood safety requirements. Instead, they validate safety by using FDA specified validation guidelines as shown in the FDA 510(k) Blood Establishment Computer Software[18].

Ruan et al. [19], [20] specify an agent-based alarm system as properties in logic, verified against formally modeled palliative care therapeutics. They created their agent-based alarms using a subset of detailed palliative workflows, checked against norms set by palliative care providers using first order LTL based modeled checker [19]. Our work captures much more than a subset of blood bank workflow; it details the vein-to-vein processes. Also, we check the properties by extracting and translating from governmental regulations from the FDA and others such as AABB.

Kristensen et al. [21] utilize a workflow management tool and LTL to verify some specified properties using so-called sweep-lines in model driven architectural design. They

focused more toward the state space explosion, but also discussed the use of checking if all properties hold in specific states. Our work differs as we focus toward ensuring blood bank regulations are checked in specific states to ensure all properties hold to meet the safety requirements utilizing the same model. Also, enforcing the same model verified into an EMR system. Devine et al. set forth approaches to verify clinical guideline properties that are linked to a model checker. Divine[22]. Their work focuses on checking the model and properties for consistency by using LTL and Promela, a model checker for clinical guidelines in Ischemic stroke prevention and management. Our work, by contrast, focuses on blood bank safety by enforcing compliance with governmental regulations and standards throughout the whole blood bank workflow.

Our work also differs from what was listed in related work by adding automated verification to ensure that practiced blood supply chains satisfy regulatory mandates. In addition, we show how EMR users can use these verified, safe workflows to provide seamless blood-related services transparent to the caregiver.

Fig. 7

A WORKFLOW GENERATED GUI

## 6. Conclusions

We have proposed and partly built what we believe to be the first system that uses generic workflow management to drive the blood collection and administration system. We have also used blood bank workflows to create a method to verify that the workflow steps satisfy the safety requirements mandated by governing authorities such as the FDA and the AABB. Our prototype shows that our methodology is sufficiently generic to model FDA mandates and AABB recommendations.

Blood safety is dynamic: Continual changes in mandates require updating the FDA regulations and AABB standards. Our system accepts these changes because the methodology is generic and, thus, can be used to specify changing safety standards and newer workflows. For example, hemovigilance started in 1994 as a means to further increase blood safety [23]. Hemovigilance attempts to track donation and transfusion processes to decrease the number of unwanted occurrences or events. We are in the process of building verification systems for Hemovigilance.

## 7. Acknowledgment

This work is sponsored in part by the Saudi Arabian Cultural Mission (SACM) and King Abdulaziz University (KAU). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, by the Government of the Kingdom of Saudi Arabia.

## References

- [1] The Patient Education Institute, Inc., "Blood Transfusion," Sep. 2013. [Online]. Available: [www.patient-education.com/bloodtransfusionpdf](http://www.patient-education.com/bloodtransfusionpdf)
- [2] "WHO | Blood safety and availability." [Online]. Available: <http://goo.gl/MBZqhf>
- [3] U.S. Food and Drug Administration, "Fatalities Reported to FDA Following Blood Collection and Transfusion: Annual Summary for Fiscal Year 2013," 2013. [Online]. Available: <http://goo.gl/kra0UK>
- [4] FDA, "Code of Federal Regulations," Apr. 2013, 00000. [Online]. Available: <http://goo.gl/m2a7fy>
- [5] C. f. B. E. a. Research, "Substantially Equivalent 510(k) Device Information - Tango Automated Blood Bank Analyzer System." [Online]. Available: <http://goo.gl/FSnryq>
- [6] J. B. Jennings, "Blood Bank Inventory Control," *Management Science*, vol. 19, no. 6, pp. 637–645, Feb. 1973. [Online]. Available: <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.19.6.637>
- [7] M. A. Cohen and W. P. Pierskalla, "Management Policies for a Regional Blood Bank," *Transfusion*, vol. 15, no. 1, pp. 58–67, Jan. 1975. [Online]. Available: <http://goo.gl/x0WDks>
- [8] Y. W. Cheng and J. M. Wilkinson, "An experience of the introduction of a blood bank automation system (Ortho AutoVue Innova) in a regional acute hospital," *Transfusion and Apheresis Science*. [Online]. Available: <http://goo.gl/9v3poy>
- [9] J. D. Roback and American Association of Blood Banks, *Technical manual*. Bethesda, Md.: American Association of Blood Banks, 2008, 00224.
- [10] Jasem Albasri, "Blood Bank Process," Mar. 2014, 00000.
- [11] "Definition of Haemovigilance : IHN - International Haemovigilance Network." [Online]. Available: <http://goo.gl/5esCiQ>
- [12] American Association of Blood Banks and Standards Program Committee, *Standards for blood banks and transfusion services*. Bethesda, Md.: American Association of Blood Banks, 2014.
- [13] B. Yu and D. Wijesekera, "Building Dialysis Workflows into EMRs," *Procedia Technology*, vol. 9, pp. 985–995, 2013. [Online]. Available: <http://goo.gl/azFi7r>
- [14] N. Hazzazi, D. Wijesekera, and S. Hindawi, "Formalizing and Verifying Workflows Used in Blood Banks," *Procedia Technology*, vol. 16, pp. 1271–1280, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2212017314003703>
- [15] Y. Foundation. (2012, 3) Yawl - user manual. [Online]. Available: <http://yawlfoundation.org/manuals/YAWLUserManual2.3.pdf>
- [16] "MAK-SYSTEM International Group," 2010. [Online]. Available: <http://www.mak-system.net/>
- [17] SCC Soft Computer, "SoftBank." [Online]. Available: <http://www.softcomputer.com/products-services/blood-services/softbank/>
- [18] "Substantially Equivalent 510(k) Device Information - 510(k) Blood Establishment Computer Software." [Online]. Available: <http://goo.gl/V93MPk>
- [19] J. Ruan and W. MacCaull, "Data-aware monitoring for healthcare workflows using formal methods," *Proceedings of the Second Workshop Knowledge Representation for Health Care KR4HC 2010, Lisbon, Portugal*, pp. 51–60, 2010.
- [20] J. Ruan, W. MacCaull, and H. Jewers, "Agent-based careflow for patient-centred palliative care," in *Electronic Healthcare*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, M. Szomszor and P. Kostkova, Eds. Springer Berlin Heidelberg, 2012, vol. 69, pp. 285–294. [Online]. Available: <http://goo.gl/4iHoXR>
- [21] Lars Michael kristensen, Yngce Lamo, Wendy MacCaull, Fazle Rabbi, and Adrian Rutle, "On Exploiting Progress for Memory-Efficient Verification of Diagrammatic Workflows," 2013. [Online]. Available: <http://cs.ioc.ee/nwpt13/abstracts-book/paper14.pdf>
- [22] A. Bottrighi, L. Giordano, G. Molino, S. Montani, P. Terenziani, and M. Torchio, "Adopting model checking techniques for clinical guidelines verification," *Artificial Intelligence in Medicine*, vol. 48, no. 1, pp. 1–19, Jan. 2010. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0933365709001365>
- [23] A. Jain and R. Kaur, "Hemovigilance and blood safety," *Asian Journal of Transfusion Science*, vol. 6, no. 2, pp. 137–138, 2012. [Online]. Available: <http://goo.gl/E4XWx7>

# Medical Image Dispersal using Enhanced Secret Sharing Threshold Scheme

Nelmiawati<sup>1</sup>, Mazleena Salleh<sup>1</sup>, and Subariah Ibrahim<sup>1</sup>

<sup>1</sup>Department of Computer Science, Universiti Teknologi Malaysia, Skudai, Johor Bahru, Malaysia

**Abstract** - Security on digital medical image storage system has become a significant concern in today's healthcare institution. Picture Archiving and Communication (PACS) which is one of current conventional storage systems for digital medical images, faced with several security issues. It fails to guarantee digital medical image survivability, confidentiality and integrity. Based on this issue this study proposed Pixel-Based Dispersal Technique on Secret Sharing Threshold Scheme. The technique combines Rabin's Information Dispersal Algorithm (IDA) and Shamir's Secret Sharing Algorithm (SSA) as a potential approach to address the issue of providing digital medical image security. Result shows that reconstructed digital medical image is exactly the same as its respective original image. Salt factor is added to enhance medical image security during image dispersal. However it requires 75% longer execution time on average, but constructs a better randomize pixels on the dispersed medical images.

**Keywords:** Rabin's Information Dispersal Algorithm (IDA), Shamir's Secret Sharing Algorithm (SSA), Pixel-Based Dispersal Scheme (PBDS), Secret Sharing Threshold Scheme

## 1 Introduction

Digital medical image comprises of pixels that represent image digitally and meta-data information to provide information about patient, imaging exercise, modality, pixel interpretation, etc. Having digitalized, it could provide a dynamic range of contrast, dynamic levels of gray, and some other manipulations to improve accuracy on image analysis during diagnose and treatment of diseases.

Storing digital medical image on a local centralized repository as offered by Picture Archiving and Communication System (PACS) does not fulfill medical image security. Disaster occurs on a local centralized storage may resulted with losing images stored in it. A recovery action from a backup system could be done only on any images that are stored before the last performed backup. Moreover, the recovery process from a backup system also often takes quite a long time until it is fully recovered. Thus, it may stop business continuity temporarily whereas it is a critical aspect to be guaranteed in any healthcare institution.

This study discusses a Pixel-Based Dispersal Scheme (PBDS), a technique used to address medical image security issues in storing medical images. The rest of this paper is organized as following: Section 2 discusses related works on secret sharing threshold scheme. Subsequently, Section 3 discusses a proposed pixel-based dispersal scheme for both salted and non-salted approaches. Furthermore, results and discussions of the scheme implementation on DICOM medical image are described in the Section 4. Next, Section 5 highlights experiment on the security services provided by the scheme. Finally, conclusion of the work is discussed in the Section 6.

## 2 Related works

Secret sharing threshold scheme is popular among distributed storage system solutions [1-3][5-8]. Rabin's Information Dispersal Algorithm (IDA) is able to encode and disperse data into several  $n$  pieces to be stored in different  $n$  sites [2]. The original data could be constructed by having at least  $m$  of its total  $n$  dispersed pieces. Thus, Rabin's IDA could tolerate absence up to  $(n - m)$  dispersed pieces for data reconstruction. This algorithm is also more efficient on its required storage size to store all of the dispersed pieces comparing with the conventional full backup [3].

Shamir's Secret Sharing Algorithm (SSA) enables construction a robust key management system that provides security and reliability. It divides a data  $D$  into  $n$  pieces  $D_1, D_2, \dots, D_n$  in such way that knowing any  $k$  or more  $D_i$  pieces makes  $D$  easily computable, but knowing any  $k - 1$  or fewer  $D_i$  pieces leaves  $D$  completely undetermined [4]. Sample use case of this algorithm is for distributing a secret key  $D$  into several entities. The secret key  $D$  can be recomputed by combining  $k$  or more number of  $D_i$ . Computation of this algorithm is based on polynomial interpolation, given  $k$  points in 2-dimensional plane  $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$ , where  $x_i$  are distinct, then there is one and only one polynomial  $q(x)$  of degree  $k-1$  such that  $q(x_i) = y_i$  for all  $i$  [5].

Each piece of  $D_i$  in Shamir's SSA has exactly the same size as  $D$ , meanwhile each piece of  $F_i$  has less size than  $F$  in Rabin's IDA. Therefore, in distributing  $D$  into  $n$  pieces of  $D_i$ , total size of  $D_i$  equals to number of pieces multiplied with size of  $D$ . This fact makes Shamir's SSA is more suitable for distributing and securing a short data such as secret key in

cryptosystem into several entities compared to distribute a big data as what Rabin's IDA can do.

The use of Rabin's IDA on data dispersal for securing query processing on relational data in a cloud system has been developed [3]. The offered solution was based on Rabin's IDA by adding salt factor for improving encoding security. B+-Tree binary index was also introduced for giving an efficient relational database operation such as select, insert, update, and delete. Besides, the use of Rabin's IDA is modified by combining All-Or-Nothing Transform (AONT) with optimized Cauchy Reed-Solomon (CRS).

Secret sharing threshold scheme was also used for providing cloud data storage system [6], [7]. Similar usage was introduced by Jigsaw secure distributed file system to securely store and retrieve files on large scale networks [1]. The Jigsaw approach is to provide a security service to the stored data by applying recursive Rabin's IDA and layered encryption. Each piece of dispersed file is sent into different distributed storages, and then these pieces are dispersed again with encoding based on a hashed-key chain algorithm derived from previous encoding key. This solution also provides a reasonable level of plausible deniability to provide users privacy and anonymity.

Both Shamir's SSA and Rabin's IDA have been implemented and tested on multi-provider clouds in securing medical records [8]. Results show several performance experiments that Rabin's IDA requires low computation.

However, none of related works found to use secret sharing scheme on securing digital medical image storage. This study introduces a pixel-based dispersal scheme on secret sharing threshold scheme by combining Rabin's IDA and Shamir's SSA to address medical image storing security issues.

### 3 Pixel-based dispersal scheme (PBDS)

PBDS is an enhance of secret sharing threshold scheme by combining Rabin's IDA and Shamir's SSA to disperse DICOM digital medical image into distributed storages for providing security services in storing medical image. Rabin's IDA is used to manipulate pixels of medical image while Shamir's SSA is used to manipulate secret key that used in the Rabin's IDA. Pixels are extracted, encoded, and dispersed into distributed storages as well as secret key. In order to reconstruct the original image, it only requires subset of dispersed files which at least equals to threshold number defined during dispersal.

PBDS has been developed in two approaches: non-salted and salted. Both have the same basic functionality on dispersing and reconstructing an image. Different from non-salted approach, the salted approach has a salt factor that produces more random pixels on the dispersed files in order to enhance the security. The following section explains more about these basic functionalities.

### 3.1 Image dispersal

Image dispersal function started by constructing a secret key. Two rules are required for the construction of the secret key, and they are as following: (a) length of secret key equals to number of servers in distributed storage, and (b) each byte of the secret key has to be unique each other.

A Vandermonde matrix secret  $C$  is generated based on the previously generated secret key for further Rabin's IDA computation. Number of rows  $n$  in the matrix equals to number of targeted distributed PACS servers to store each of produced dispersed images, and number of column  $m$  in the matrix equals to threshold minimum number of required PACS servers for retrieving the dispersed images successfully in order to reconstruct back the original image.

Galois Field  $GF(2^8)$  computation is applied in PBDS since the allocated size for each band on pixels of DICOM medical image is always 1 byte. The objective is to make sure the manipulation of a pixel is still resulted with a valid pixel value. As an example:  $n = 4$ , and  $m = 3$ , by following the given rule, randomly generated secret key has to be 4 byte length, such as:  $\{0 \times 01, 0 \times 07, 0 \times 0A, 0 \times 0F\}$ . Following Vandermonde matrix secret  $C$  is generated based on that secret key:

$$C = \begin{bmatrix} 01^0 & 01^1 & 01^2 \\ 07^0 & 07^1 & 07^2 \\ 0A^0 & 0A^1 & 0A^2 \\ 0F^0 & 0F^1 & 0F^2 \end{bmatrix} = \begin{bmatrix} 01 & 01 & 01 \\ 01 & 07 & 15 \\ 01 & 0A & 44 \\ 01 & 0F & 55 \end{bmatrix}$$

Pixels and header from DICOM medical image file are extracted after generating the secret key. Sequences of pixels are grouped into blocks, then each of them are transformed into a matrix  $D$ . Number of pixels in one block equals to  $m$  which is the number of threshold dispersed image to reconstruct the original image. Each pixel consists of one or more bands. As an example, one band for monochrome and three bands (red, green, and blue) for RGB.

First pixel in the block is become the first row in the matrix  $D$ , the second pixel is become the second row; thus  $m^{\text{th}}$  pixel is become the  $m^{\text{th}}$  row. Meanwhile, the first band of the first pixel is become element  $e_{1,1}$  in the matrix  $D$ , the second band of the first pixel is become element  $e_{1,2}$  and the  $x^{\text{th}}$  band of the first pixel is become element  $e_{1,x}$ . Taking an example of a sequence of first five pixels for an RGB DICOM image =  $\{11D1F2, F1A271, 10017A, F0112D, 31F2B1\}$ . Suppose  $m = 3$ , then the first 3 pixels  $\{11D1F2, F1A271, 10017A\}$ , are considered as one block. It could be transformed into a matrix  $D$  as following:

$$D = \begin{bmatrix} 11 & D1 & F2 \\ F1 & A2 & 71 \\ 10 & 01 & 7A \end{bmatrix}$$

A salt factor is added into a matrix  $D$  for salted PBDS before the encoding process. It is created by using

Pseudorandom Number Generator (PRNG) function  $fs$  with seed  $ss$  equals to  $\sum a_i$  of the Vandermonde matrix secret.

Given a Vandermonde matrix secret  $C$ . A second column  $C_{:,2}$  equals to the secret key. All of the elements in  $C_{:,2}$  are summed up to create a secret seed  $ss$ . Suppose  $fs(ss)$  is a pseudo random generator function with  $ss$  as its seed. Before encode and dispersal processes, each row in  $D$  is added with different salt factor  $fs.nextRandom$ . The salt factor is added to each row to randomize each pixel of medical image. It is possible to remove characteristics of corresponding original medical image in every dispersed files generated completely. Equation (1), (2), and (3) describes formula to generate salt factor.

$$ss = GF(C_{1,2} + C_{2,2} + \dots + C_{n,2}) \quad (1)$$

$$fs(ss); fs = PRNG, ss = \text{secret seed} \quad (2)$$

$$D = \begin{bmatrix} a & b & c \\ 1 & 2 & 3 \\ \dots & \dots & \dots \\ x & y & z \end{bmatrix} + \begin{bmatrix} fs.nextRandom \\ fs.nextRandom \\ \dots \\ fs.nextRandom \end{bmatrix} \quad (3)$$

The encoded matrix  $E$  is computed as a result of multiplication between matrix secret  $C$  with matrix  $D$  as shown in Equation 4. Each row in the matrix  $E$  becomes one new pixel to be written into its respective distributed server. Number of distributed servers equals to number of row in the matrix  $E$  as in Equation (4). The encoding process is applied to all of the blocks sequentially until the last block of pixels. Padding pixels will be added in the last block if number of remaining pixel is less than  $m$ .

$$E = C \times D \quad (4)$$

The secret key  $\{0 \times 01, 0 \times 07, 0 \times 0A, 0 \times 0F\}$  is split up by using Shamir's SSA into several shared secret keys. Number of shared secret keys has to be same as the number of the distributed servers as well. The generated shared secret keys are stored at the header tag on each of generated new dispersed DICOM image files.

All of DICOM tags header in the original image, except  $0 \times 7FE$  (pixel data) need to be copied into each of new DICOM files to be dispersed. In addition, following proprietary PBDS, private tags need to be added into each of new DICOM images:

1. Tag ( $0 \times 0029, 0 \times 0010$ ) is any string value to indicate the creator name.
2. Tag ( $0 \times 0029, 0 \times 1000$ ) is dispersal index value, index for each of new dispersed DICOM file. The index value is in the range of  $0 \leq i < n$  number of targeted servers.
3. Tag ( $0 \times 0029, 0 \times 1001$ ) is a shared secret key value.
4. Tag ( $0 \times 0029, 0 \times 1010$ ) is original medical image width value.
5. Tag ( $0 \times 0029, 0 \times 1011$ ) is original medical image height value.

Finally, a new generated dispersed DICOM files are ready to be stored in their respective distributed servers.

### 3.2 Image reconstruction

Image reconstruction process performs original image reconstruction from its dispersed images.  $m$  numbers of dispersed images are sufficient to reconstruct the original image. The shared secret keys are extracted from retrieved dispersed images to reconstruct the secret key by using Shamir's SSA. Next, it is used to generate matrix secret key  $C$  that will be used during dispersed images decoding process into its original image by using Rabin's IDA.

The processes are started by pixel decoding. It involves  $m$  numbers of retrieved dispersed images. Matrix  $E^*$  is generated from those  $m$  numbers retrieved from the dispersed files. One pixel from the first dispersed image will be the first row of the matrix  $E^*$ , second row matrix  $E^*$  contains second dispersed image and etc; thus  $m^{\text{th}}$  pixel of dispersed image lies at the  $m^{\text{th}}$  row of the matrix  $E^*$ .

The order of the retrieved dispersed images is important. It follows the order during dispersed process. Private tag ( $0 \times 0029, 0 \times 1000$ ) is added into the dispersed images is used to preserve this order. By following the correct order matrix  $E^*$  will be the sub-matrix of  $E$  with less number of rows. Next process is shown by Equation (5) whereby matrix  $D$  is calculated by multiplying  $C^{*-1}$ , that is the inverse of sub-matrix secret  $C^*$ , with  $E^*$  [3]. Matrix  $C^*$  is a sub-matrix of  $C$  with some of its rows are not existing, similar to matrix  $E^*$ . Matrix  $D$  is a decode result that will reconstruct original image. Row dimension of matrix  $D$  is equal to  $m$  and column equals to pixel's band number.

$$D = C^{*-1} \times E^* \quad (5)$$

Matrix  $D$  needs to be processed further in salted PBDS approach. Salt factor that is added on dispersal process needs to be subtracted during reconstruction process. Finally, pixels are ready to be formed from matrix  $D$  to reconstruct the original medical image.

## 4 Results and discussions

The experiments were focused on: (1) dispersal and reconstruct quality for both approaches; (2) quality of computing performance in running PBDS on salted and non-salted approach; (3) storage requirement needed to run PBDS.

### 4.1 Dispersal and reconstruction quality

A DICOM medical image is used for validating for both PBDS approaches: salted and non-salted. Experiments were conducted with four distributed servers while the threshold equals to three, which means four dispersed files will be generated and stored into four distributed servers, at least any three pieces of the dispersed files are sufficient to reconstruct the original image.

Results in the Fig. 1 prove that the original image and the reconstructed image are exactly identical in pixel to pixel. Peak Signal-to-Noise Ratio (PSNR) value of both images equals to infinity which means that both are exactly the same.

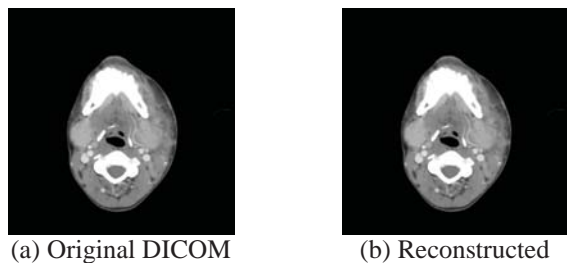


Fig. 1 Original and reconstructed image

4.1.1 Salted

Fig. 2 shows four dispersed files as results of dispersing original DICOM image (Fig. 1) by using salted PBDS. It shows that the generated dispersed files with salted approach do not reveal any characteristics of the original medical image.

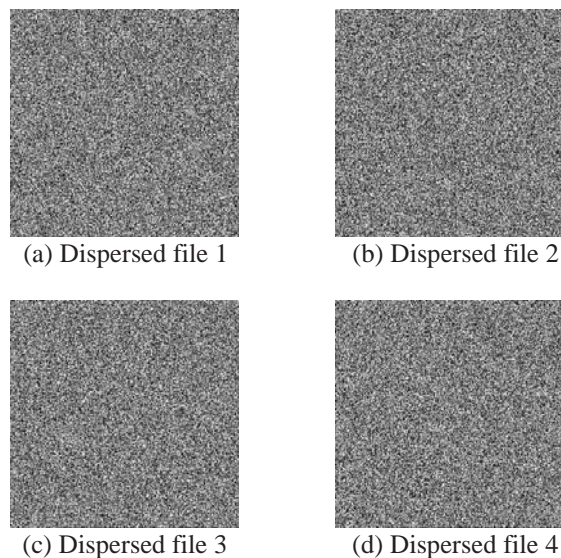


Fig. 2 Dispersed salted images of Figure 1 original DICOM

4.1.2 Non-salted

Fig. 3 shows four dispersed files as results of dispersing original DICOM image (Fig. 1) by using non-salted PBDS. Same patterns as the original medical image are shown. The scramble parts only cover the middle area of the original image in dispersed files. The other surrounding parts remain black. Black color represents zero value. When all of its components are zero produces zero value as well on matrix multiplication. Thus, dispersed files generated with non-salted approach have less confidentiality since some characteristics of the corresponding original medical image can be revealed in several cases. However, both approaches, salted and non-salted are able to reconstruct the exact original medical image.

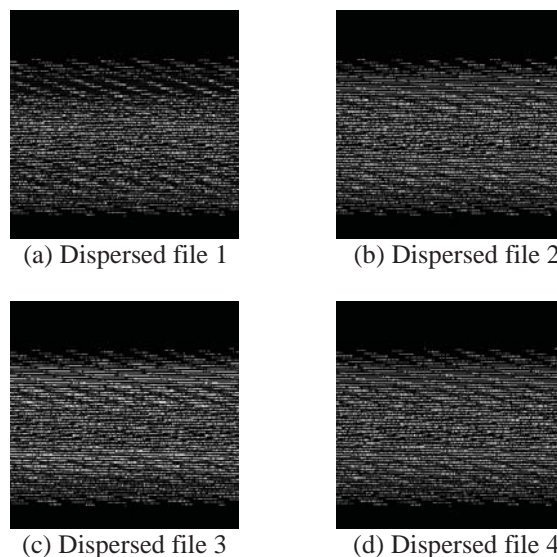


Fig. 3 Dispersed non-salted images of Figure 1 original DICOM

4.2 Performances

Performance test was performed to compare execution time during dispersal and reconstruction medical image for both approaches; salted and non-salted is measured. There are 360 DICOM medical images were tested on each of approaches. Fig. 4 and Fig. 5 show comparison of execution time for images dispersal and reconstruction respectively, both applied to salted and non-salted approaches.

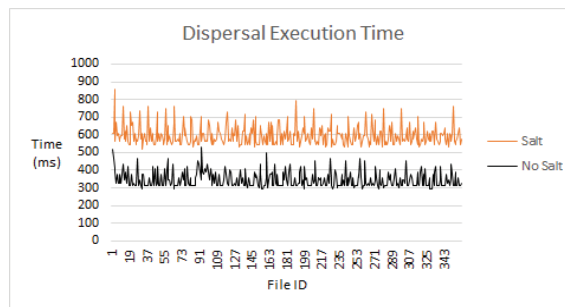


Fig. 4 Dispersal execution time



Fig. 5 Reconstruct execution time

Fig. 4 shows the execution time required for dispersal with salt approach is about 73% longer than the one with non-

salted approach. Fig. 5 shows about the same result for reconstruction function. Reconstruction with salted approach is about 76% longer than the one with non-salted approach. It could be explained that extra time was needed for salt factor computation.

On the other hand, CPU and memory usages to process both approaches are approximately the same as shown by Fig. 6 till Fig. 13. CPU usage on disperse process for salted and non-salted approach is about 20 to 25%, while memory usage is about 5 to 10MB. Similar result is given during reconstruction process. It could be concluded that salted computation is light enough thus it does not require a lot of CPU and memory usages.



Fig. 6 CPU usage for disperse non-salted



Fig. 7 CPU usage for disperse salted

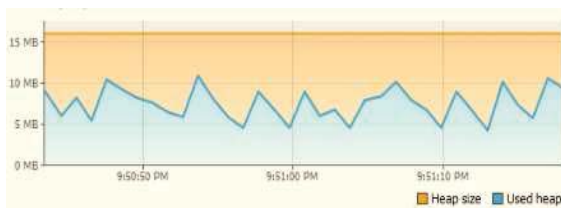


Fig. 8 Memory usage for disperse non-salted

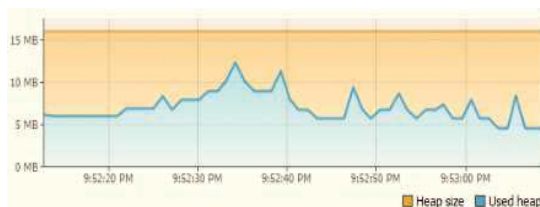


Fig. 9 Memory usage for disperse salted

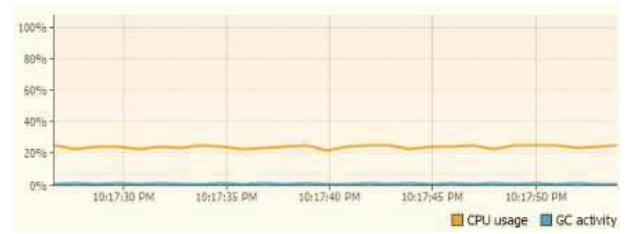


Fig. 10 CPU usage for reconstruct non-salted



Fig. 11 CPU usage for reconstruct salted

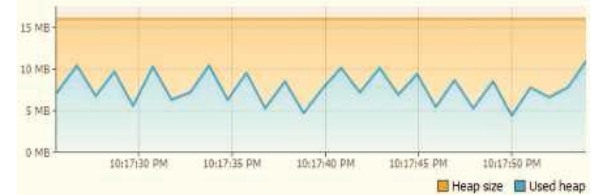


Fig. 12 Memory usage for reconstruct non-salted

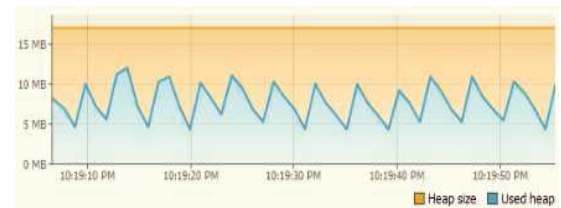


Fig. 13 Memory usage for reconstruct salted

### 4.3 Storage requirement

PBDS provides an economical storage requirement. Suppose, a system have  $b$  backup servers, total space for storing medical images with size  $F$  pixels is given by Equation (6).

$$[b \times F] \quad (6)$$

Through the proposed scheme; both salted and non-salted, suppose number of targeted distributed servers  $n = b$ , and  $m =$  threshold number where  $m < b$ , the total space for storing medical images with size  $F$  pixels is given by Equation (7).

$$\left[ \frac{n}{m} \times F \right] \equiv \left[ \frac{b}{m} \times F \right] \quad (7)$$

Since  $n = b$  and  $0 < m < b$ , then:

$$\left[ \frac{b}{m} \times F \right] < [b \times F] \quad (8)$$



Thus, Equation 8 shows efficient storage capability on the proposed solution compared to conventional backup system.

## 5 Experiment on security services

PBDS provides several security services to solve issues arise in storing medical image. It ensures survivability when no more than  $n - m$  servers are faulty. Besides, it ensures confidentiality and integrity when wrong shared secret key is used or corrupted dispersed files used to reconstruct medical image.

### 5.1 Survivability

Medical image survivability exists on both PBDS approaches. The original medical image could be fully reconstructed as long as number of dispersed files are not less than a threshold number  $m$ . Suppose four dispersed files generated to be dispersed into four distributed servers with threshold number is three, then down in any of one of the servers could be tolerate to fully reconstruct original medical image.

The conducted test proves that pixels on the reconstruct image are exactly the same as the ones at the original image. PSNR between them equals to infinity that means exactly the same.

Each piece of dispersed files are produced during dispersal as shown in Fig. 15. Small PSNR value between them shows their differences are significant. PSNR within dispersed image in Figure 15 is given by Table 1.

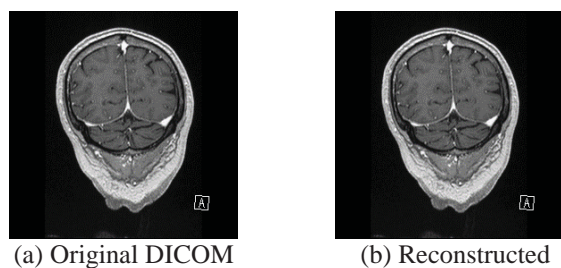


Fig. 14 Image survivability

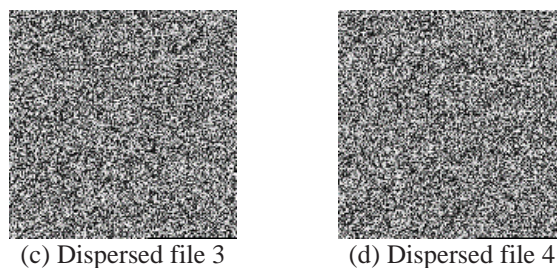
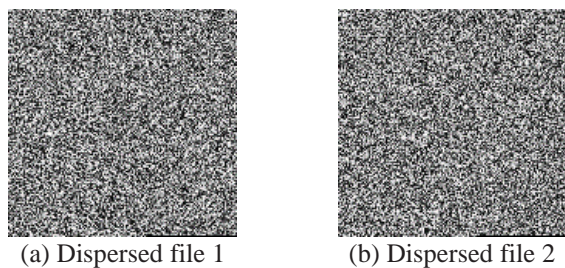


Fig. 15 Four dispersed files of Figure 14 original DICOM

Table 1 PSNR (dB) of the dispersed files

	A	B	C	D
A	infinite	7.8	7.77	7.78
B	7.8	infinite	7.71	7.77
C	7.77	7.71	infinite	7.73
D	7.78	7.77	7.73	infinite

### 5.2 Confidentiality

Medical image confidentiality relates to protection from a disclosure by preventing unauthorized entities to get access into it. Suppose a third party steals two of three dispersed images successfully. The third party also has knowledge to interpret the stolen dispersed images perfectly, and might guess the pixels value to construct the third dispersed image correctly. However, the third party is not able to get the correct shared secret key that is kept together with the third dispersed image. Therefore, the third party will not be able to reconstruct the original image. Fig. 16(a) shows original medical image while Fig. 16(b) is the reconstructed image when a wrong secret key was used during reconstruction.

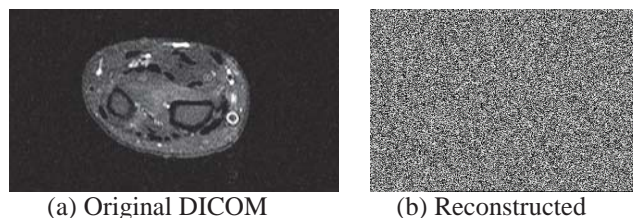


Fig. 16 Image confidentiality

### 5.3 Integrity

In the integrity testing, one of the dispersed files used for original image reconstruction, has its pixels partially corrupted. It shows that the reconstructed image is also corrupted as much as  $t \times$  number of corrupted bytes, where  $t$  equals to the threshold number. On the other hand, if the corruption occurred in the header of the dispersed image, then reconstruction of the original image will definitely failed. It could be happened due to DICOM header contains medical image information to interpret each of byte in medical image. Fig. 17 shows an original image and its reconstructed image when one of the dispersed file use for reconstruction was corrupted, meanwhile Fig. 18 shows the dispersed files used for the reconstruction, given that one of them is corrupted.

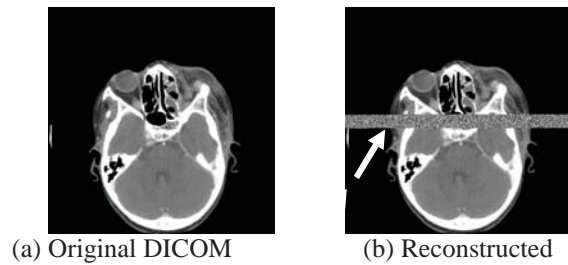


Fig. 17 Image integrity

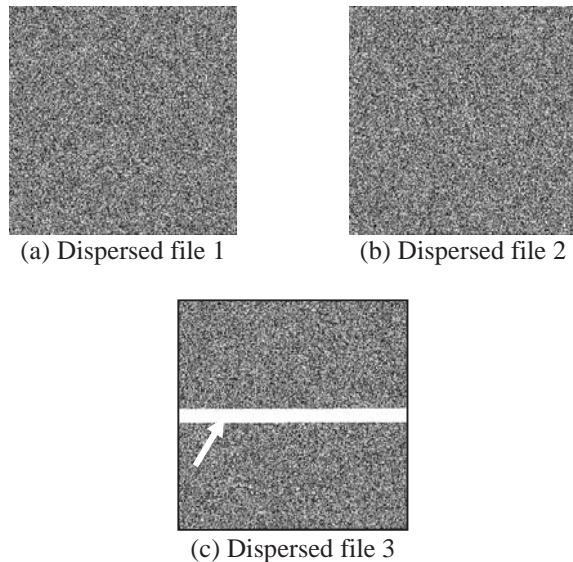


Fig. 18 Three dispersed files of Figure 17 original DICOM

## 6 Conclusions

Digital medical image is very important in healthcare institution. Security issue is a big concern in managing it. Pixel-based dispersal scheme has been developed in addressing security issues in storing medical image, following two approaches: non-salted and salted. Salted PBDS produces more random results on the dispersed images compared to non-salted approach. Even though salted PBDS requires longer execution time, it does not require significant extra CPU and memory usages. In fact, storage requirement for both of these approaches are the same. Therefore, it could be concluded that salted PBDS gives a better quality on providing security services with accepted drawback on the resource utilization. Further study could be done on leveraging implementation of PBDS in a cloud computing environment.

## 7 Acknowledgments

In completing this study, it is supported by Ministry of Education (MOE), Malaysia and UTM under Vote No. (4L108).

## 8 References

- [1] J. Bian and R. Seker, "The Jigsaw Secure Distributed File System," *Computers & Electrical Engineering*, vol. XXXIX, no. 4, p. 1142–1152, 2013.
- [2] M. O. Rabin, "Efficient Dispersal of Information for Security, Load Balancing, and Fault Tolerance," *Journal of the ACM*, vol. 36, no. 2, pp. 335-348, April 1989.
- [3] S. Wang, D. Agrawal and A. E. Abbadi, "A Compression Framework for Secure Query Processing on Relational Data in the Cloud," in *Proceeding SDM'11 Proceedings of the 8th VLDB international conference on Secure data management*, Seattle, 2011.
- [4] P. V. M., S. Han and C. E., "Fingerprinted Secret Sharing Steganography for Robustness Against Image Cropping Attacks," in *International Conference on Industrial Informatics*, 2005.
- [5] A. Shamir, "How to Share a Secret," *Magazine Communications of the ACM*, vol. 22, no. 11, pp. 612-613, November 1979.
- [6] D. Slamanig and C. Hanser, "On Cloud Storage and the Cloud of Clouds Approach," London, 2012.
- [7] H. Lahkar and M. C. R, "Towards High Security and Fault Tolerant Dispersed Storage System with Optimized Information Dispersal Algorithm," *International Journal of Advance Research in Computer Science*, vol. 5, no. 6, pp. 286-290, 2014.
- [8] T. Ermakova and B. Fabian, "Secret Sharing for Health Data in Multi-Provider Clouds," in *Conference on Business Informatics (CBI)*, Vienna, 2013.

# A Diagrammatic Notation for Care Coordination

W. Zachary, PhD<sup>1</sup>, and R. Maulitz, MD., PhD<sup>1</sup>

<sup>1</sup>Starship Health Technologies, Plymouth Meeting, Pennsylvania, USA

**Abstract** - Coordination of care has emerged as a common issue in the prevention of medical errors across the entire population, improvement of outcomes for chronic illness in elderly patients, and lowering the cost of care. However, there are currently no methods or analytical tools for decomposing, formalizing, and representing the processes that unfold in individual episodes or arcs of care coordination. A new method to meet this need, called Coordination Process Diagramming or CPD, is presented and demonstrated in the context of two clinical examples of increasing complexity. The CPD method has broad applicability in analyses and simulation modeling of care coordination processes, and in the development of informatics tools to support and facilitate care coordination itself.

**Keywords:** care coordination, communication networks, clinical communications, health systems analysis, coordination process diagramming

## 1 Introduction

Care coordination is recognized as central to accomplishing a range of health care goals, including improving outcomes and transitions of care, enabling accountability and achieving systemic financial sustainability, and reducing medical errors. At the policy level, efforts to eliminate medical errors [3,4], manage chronic illness in an aging population [5], and reduce costs while increasing effectiveness of care [2,11] are also being addressed, at least in part, by focusing on care coordination. However, while care coordination is increasingly seen as a significant process in health systems, research and policy analyses have not proceeded from a grounded empirical understanding, but rather from highly aggregate data (e.g., relying on aggregate data on re-admission rates) or detailed examination specific clinical vignettes.

A recent example of the latter can be found in Press [6]. He describes and analyzes a specific, complex, and protracted case in terms of the (large number of) clinical communications, through which coordination occurred. Here, we argue that this level of analysis – the component interactions that comprise a single care coordination episode or arc – is the only level at which specific failings, successes, features, and patterns of care coordination can be identified, analyzed, and ultimately used to develop an evidence base from which to address health care and health policy concerns. What is lacking, however, is more formalized language in

which to frame and express this discussion. Without a common formalized language that can be used to describe, compare, and analyze individual care coordination arcs, the discussion remains at the level of anecdotes compared through armchair analysis. In this paper, we describe and give an example of such a formal language for describing and analyzing care coordination.

## 2 Basic Features of Coordination Processes

Different informatics fields (e.g., information systems analysis, software design, organizational design) have developed and adopted a broad range of formal methods to enable structured description, comparison, and analysis. The forms of these, however, have ranged from highly mathematical to purely diagrammatical. Which approach might be best for health care informatics? A cue can be taken from [6] who interestingly used a network diagram to map out all the communications in an 80-day care coordination episode. Social network representations (e.g., [9, 10]) have some prior use in research into clinical communications [9,7,1] and do provide part of a framework for representing how coordinated care proceeds. However, the basic social network diagram, in which individuals are depicted as nodes and communication relationships are depicted as lines between nodes, is insufficient by itself. Two individuals may communicate at many points in time and for different purposes during a care coordination arc. Each of these communications needs to be represented separately. This is what Press did, but the resulting diagram showing many parallel links between the same two persons (nodes) is visually very complex. Perhaps this is why a graphical animation was added (in the on-line publication) -- to show where individual communications occur in the overall sequence. However, animation is not an effective analytical tool (even though it is a valuable form of scientific visualization). Rather, the temporal dimension of individual communications must be explicitly represented in the communication network diagram in some other way.

Another feature of care coordination is that communication processes often cross organizational boundaries, both minimal (e.g., different departments in a given facility) and more significant (e.g., different facilities, different stand-alone practice groups). The major boundaries may also represent portions of the overall network that cannot easily share data (because they use differing electronic health record [EHR] systems, and/or are in different states, or even countries with different data sharing laws and policies) relying on existing

social networks of relationships among clinicians to chart these crossings. Thus, not only must organizational boundaries be an explicit part of the formal representation, specific data repositories within a boundary may also be important. Providers in a care coordination process who are within a given organizational boundary may not only be able to share patient data that is relevant to the communication more easily, they often can communicate indirectly through EHRs via tasking note mechanisms.

The final core feature of care coordination is that while the bulk of the communications are between and among clinicians and other health care personnel, the patient and patient's network can also play a key part of the process. The patient's network in this case could include a spouse or significant other party who can provide information about the patient or patient's personal or family history when the patient is unavailable or unable to communicate. Holders of durable power of attorney can also be part of care coordination discussions with parts of the clinical team. We generally do not include direct clinical encounters or procedures as part of the care coordination process, although incidents in such encounters may become part of the care coordination process and hence included in the description of the arc.

### 3 Basic Coordination Process Diagramming

Coordination Process Diagramming or CPD is an analytical technique for decomposing and diagramming a specific arc of care coordination. From the above discussion, the five basic rules of the CPD are defined as follows:

1. Communication network comprised of nodes and links. A care coordination arc is broken down into individual communications between/among people or between people and data repositories (e.g., EHRs from which data are retrieved or through which indirect communication occurs via tasking mechanisms). Each person is denoted as a circle or dot, and each data repository is denoted as a box; together circles and boxes are termed 'nodes'. A specific communication is denoted as a link or line connecting two circles (person-to-person) or a circle and a box (human-to-repository). If one node initiates the communication (e.g., Dr. Jones phones Dr. Smith), then the link is an arrow from Dr. Jones' node to Dr. Smith's node). If Dr. Smith retrieves a tasking communication from an EHR, then that is depicted as an arrow from the EHR node to Dr. Smith's node. If Dr. Jones emails both Dr. Smith and Dr. Fernandez with the same email, the link is two arrows with a common origin in Dr. Jones' node and separate ends in Dr. Smith's and Dr. Martinez' nodes. Each separate communication between two nodes is represented by a separate link.

2. Ordering of nodes based on time. Unlike a conventional social network diagram, the nodes are positioned in the diagram in the order in which they enter the care coordination arc. If the care coordination diagram is constructed left-to-right, the left-most node will be the clinician who initiates the

first communication, and the person/node to whom that communication is directed will be placed to the right of that node. If two nodes enter at the same time, they can be placed above/below each other.

3. Temporality annotations on each link. Every node in the communication network has an annotation associated with it in the form  $t(i)$ , where  $i$  refers to the order of that link in the overall coordination arc. Thus, the link denoting the first communication is annotated  $t(1)$ , the second as  $t(2)$ , etc.

4. Explicit depiction of organizational boundaries via layers. The direction in which nodes in a CPD are temporally ordered is called the main axis (e.g., typically either left-right or top-down). The perpendicular axis is the secondary axis, and is used to separate nodes by organization. All nodes that belong in the same organization are placed in the same layer, with layers separating organizations denoted by dashed lines. The organizational name is shown in the originating end of the main axis. These layers make clear where organizationally each participant in the care coordination arc works.

5. Separate layer for patient, family, and community contacts. A special case of rule 4 concerns nodes that represent the patient or members of the patient's network (e.g., spouse, other family member, holder of durable power-of-attorney). All of these nodes will be placed (in order of temporal entry, according to Rule 1) in a special layer that is above the initiating node's organization on a secondary axis for a horizontal main axis (or to the left of it for a vertical main axis). This layer separates all the non-clinical nodes and communications from those that involve the patient or the patient's network.

A basic coordination process description of a care coordination arc then consists of a coordination process diagram showing all the nodes involved, with the nodes ordered in the main axis according to the sequence in which they become involved in the arc. Each communication between a node is shown as a directed link with a time annotation indicating the sequence number of that communication in the overall arc. The nodes are further grouped in the secondary axis into layers representing the organization to which they belong, with the patient and patient-related non-clinical nodes occupying the first layer on the secondary access. There is also an optional separate table that defines the chronological time at which each communication occurred. These chronological times can be either absolute (i.e., tied to calendar and clock times) or relative (e.g., to the time of the first communication).

Figure 1 shows a simple example of CPD, based on the following narrative description.

Mr. X, a 77 year old male with a past medical history of hypertension and obesity, presents with new onset left lower leg pain. The patient reports having low leg pain for the past two weeks. He reports having driven from South Carolina to Maryland two weeks previously, noticing the swelling shortly after that trip. He denies chest pain and shortness of breath.

After seeing his Primary Care Physician, Dr. A in Dr. A's primary care office, Dr. admits him to Community Hospital A. There, Dr. A requests a duplex ultrasound of his left leg, which proves positive for deep vein thrombosis. Dr. A then starts Mr. X on anticoagulation (Lovenox and Coumadin) after EHR confirmation of a prior negative colonoscopy report. On day 2, Dr. A, on rounds, discusses the following: with nutritionist B on diet related to Coumadin use, with nurse C to educate the patient on home Lovenox injections, and with case manager D on obtaining prior authorization for

the patient to receive home Lovenox injections, as well as to have a visiting nurse E scheduled to see the patient to ensure proper delivery of the injections. At discharge, the in-patient resident F taking care of Mr. X communicates with Dr. A, providing details of the Coumadin dosing: Coumadin 5mg daily, with home Lovenox injections subcutaneously 100 mg twice daily. The patient is discharged to home, obtaining the scheduled INR 3 days later to assess adequacy of anticoagulation.

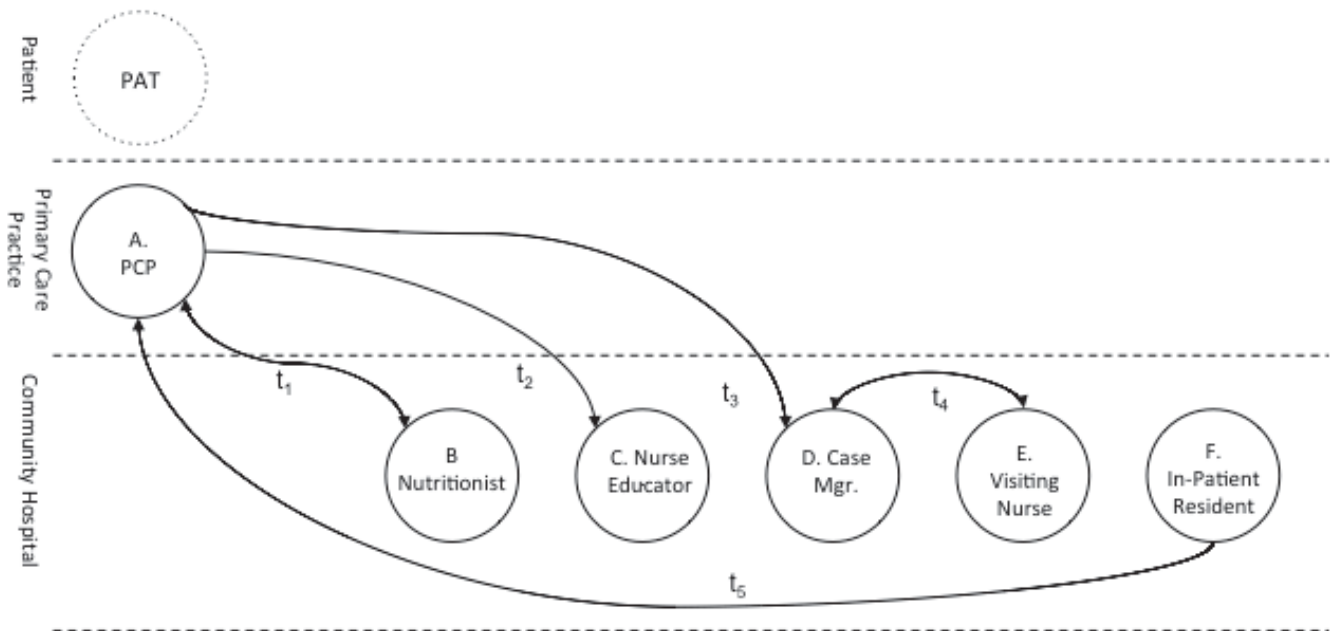


Figure 1. Simple Example of a Coordination Process Diagram

#### 4 The Communication Process Content Table

While Figure 1 shows a great deal of information about the basic structure of this simple care coordination episode, it should also be immediately clear that the *content* of the arc is missing. That information is contained in the individual communications themselves, and requires another level of description. This second (content) level is expressed in a tabular form called the Communications Process Content Table, or CPCT. The basic unit of the CPCT is an individual link in the Coordination Process Diagram. For each link, three types of information about the communication it represents are broken up in the CPCT. We note that the term ego refers to the person who initiates the communication and term alter(s) refers to the person(s) engaged by ego in the communication; this terminology is consistent with social network methods [e.g., [11,12]. In several cases below, the content description is based on sets of descriptive categories developed by the authors in prior methodological research on methods to document care coordination communications [12].

The first type of content is information about the communication itself. This includes the:

- *functional intent* of the communication, using a set of intent categories developed in [12]; the idea here is to categorize why ego was communicating with alter; and
- *semantic content* of the communication, summarized in free text (as in the vignette above captured in Figure 1). Such information could later be analyzed using text analysis methods such as latent semantic analysis [13].

The second type of content is information about the Link itself. This includes the:

- *specific individuals involved* (inherited directly from the CP diagram), and the directionality of the communication involved. Generally it is either directional (from ego to alter, typical for a declarative communication) or bidirectional (typical of back-and-forth dialogs, such as when two clinicians are trying to converge their understanding of the current situation);

- *roles of the individuals involved*, categorized in broad functional terms -- primary care physicians (PCP), secondary- or tertiary-care specialist (SP), trainee, i.e., resident or intern (TR), nurse (N), technician (TN) etc. Such roles de facto constrain the functions in which each party in the communication can serve, and provide a means to explore the data in structural as well as individual levels of analysis; and
- *modality of the communication*, coded as either face-to-face interactions, electronically-mediated direct communications (phone calls, e-mails, SMS or other messaging), and indirect methods, such as 'tasking' annotations in EHRs or notes on visually-shared artifacts, e.g., via white-boards or computer monitors.

The third type of content involves the information on patient context that is communicated or that accompanies the communication.

It includes the:

- *problem list items, medications, planned or completed studies and results* (all as occur in the communication), potentially grouped into categories to capture commonly occurring information, such as types of pharmaceuticals, types of tests, types of procedures, etc.
- *name or identifier of the patient's PCP or attending physician*. Several researchers [e.g., 8,6] argue that the PCP in particular is in a unique position to add value to downstream care coordination communications, yet it is unclear whether any information about this individual is appropriately propagated downstream in arcs;
- *patient social context information*, such as living status and availability of social support, and
- *patient gender, age, race and minority status, and insurance status*. Among other things, such data can allow the effects of health disparity effects to be recorded and analyzed.

The overall form of the CPCT is a table in which each row (or column) represents a specific link in the CP diagram. The order of the links should reflect the temporal order of the communication in the overall process. Thus link t(1) is first, link t(2) is second, and so on. Each additional column (or row, if using an inverted form) then is used to record information of each type in each category described above. This layout allows an easy inspection of, for example, how important patient context information is appropriately propagated over time, or gets lost (resulting, for example, in re-doing tests that had previously been done).

## 5 Advanced CPD Features

Care coordination in the real world can become complex and

can occasionally exhibit dysfunctional activities. In this subsection, we discuss a few final features of coordination process diagramming intended to address these issues. Each is briefly introduced below.

Transitions of Care (ToC). One kind of complexity not discussed in 3 above is the transition of care. An oft-stated policy goal for care coordination is to minimize the number of transitions of care, and to increase the time between such transitions of care (e.g., avoiding hospital re-admissions within 30 days of discharge). Longer care coordination arcs may include one or even several such transitions, so a symbol has been included in the method to denote them. The symbol is a jagged vertical line with annotations on each side indicating the ending and starting state of the patient. State in this sense can be "ambulatory care", or an in-patient location (e.g., "hospital A", or "rehabilitation/nursing facility B" or "hospice C"). The ToCs are placed on the bottom layer of the diagram, at a point in time corresponding to the transition on the main axis of the diagram. An optional time annotation, in the same form used for links, can be placed on the transition to situate it more precisely in the temporal scale of the arc.

Failed communications. Unfortunately, not all attempts at communication succeed. Anecdotal accounts of care coordination arcs often include points where one node repeatedly attempts to contact another, but is never able to do so. To indicate such failed attempts, a failed communication symbol is used. This symbol is simply a dashed grey-shaded version of a one-way communication link. The time annotation indicates the time the communication was first attempted; a second time annotation can indicate the time at which the effort was ended.

Indirect (EHR-mediated) communications. Many EHR systems allow clinicians to leave tasking or directive notes for other clinicians. These are a form of indirect communications because the communication may involve a substantial latency between the time the note is created and the time it is received. In some cases, the note may never be successfully retrieved. For this reason, CPD breaks this link into two parts – the first part in which the initiating person creates the EHR note (denoted as a link from the node to the EHR symbol, annotated by the initiating time), and the second part in which the recipient actually receives it (denoted by a separate link from the EHR symbol to the recipient, annotated by the time of actual retrieval). This also allows for some indirect communications to become de facto failed communications.

Starting and ending conditions. For chronic illnesses, it may be difficult to pinpoint a precise beginning and end to a long care coordination episode, as multiple arcs involving different underlying problems may be on-going at the same time. Accordingly, CPG has included optional originating and ending points to be defined. The originating point (pictured as a diamond with an arrow leading the source of the initial communication link in the arc) can include an event such as test result or new complaint as the initiating point for the arc.

Its implicit time annotation is  $t(0)$ . The ending point (pictured as a final link from the recipient of the last communication link to a diamond) can include a terminating condition or state. A terminating condition could be such criteria as “no further communications on the arc after 30 days”, or “released to home care for primary care follow up” or “third readmission on this arc”. It could also be a state (e.g., death) or transition to rehab or nursing care.

Figure 2 shows a summary of all the graphical building blocks in the Coordination Process Diagramming method.





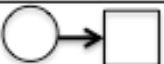
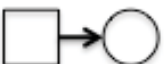





Node: A person in the care coordination process	
Node: Data repository or (EHR)	
Link: One-way communication	
Link: Two-way communication	
Time-annotation for link	$t(i)$
Node puts asynchronous task into EHR	
Node retrieves asynchronous task from EHR	
Organizational brackets for nodes in organization "name"	
Failed communication	
Transition of care (from A to B)	
Originating event for care coordination arc	
Terminating event for care coordination arc	

Figure 2. Complete Coordination Process Diagramming Notation

## 6 A More Complex Example

As a final example, we present the following narrative, which is depicted as a CPD in Figure 3.

Ms. X, a 68 year old Caucasian female with a history of diabetes, hypertension, morbid obesity and intermittent abdominal pain, presents to Dr. A (PCP) for a recurrence of her pain. She notes it is in the epigastrium (area near stomach) and seems to follow meals. Worrisomely, her rectal

exam showed occult blood. The abdominal examination is unremarkable.

She is therefore referred to Dr. B, a colorectal surgeon, for a work-up to include a colonoscopy. Dr. B performs an initial examination and notes “no GI complaints” and “no distension” on physical. Dr. B sets Ms. X up for a colonoscopy, to be performed in the operating room under anesthesia because of the convoluted large intestine. Dr. B sends a task detailing the plan going forward to Dr. C, the head of Dr. B’s group. Dr. C then forwards this task to Dr. A to complete the circle and make him (Dr. A) aware of Ms. X’s current status.

Four days later Ms. X returns to Dr. A, who is poised for the encounter because of the completed background communication. Hearing hyperactive bowel sounds and a continued complaint of abdominal discomfort. Dr. A prescribes hyoscyamine, a drug that when taken by mouth or in sublingual form reduces intestinal motility. Dr. A urges the patient to continue her work-up with Dr. C.

A week to week-and-a-half after these events, the patient presents to the emergency room with worsening pain. ER Attending MD Dr. F calls in Dr. B to evaluate the patient, admitted her for observation, and orders cardiac-clearance from Dr. D, a cardiologist. At this point Dr. A’s service is contacted by Dr. B; the respondent is Dr. E, a resident (trainee), on the in-patient team, who sends an EHR task to Dr. A detailing the admission. Dr. A reads it later that day.

Dr. E’s tasking note states that “pt X was admitted to University Hospital 4/22-4/23 for abdominal pain/nausea/vomiting and acute cholecystitis, and was discharged after IV fluid rehydration and pain medications.” The task goes on to say “Ms. X was recalled and readmitted on 4/24 after both blood cultures [were] positive for gram negative rods.”

The patient is now placed on vigorous hydration and antibiotics by the two (primary care and surgery) in-patient teams. As of April 26<sup>th</sup>, no one from these in-patient teams has contacted Dr. A with an update on Ms. X’s status. On this date, Dr. A therefore calls Ms. X’s home in hopes of getting updated information from her or her husband. There is no answer. Dr. A leaves a message requesting a return telephone call. An hour later, Mr. X called on behalf of his wife and provided a final update. Though still in the hospital, Mr. X states that his wife is improving on antibiotics after having a colonoscopy that showed diverticulitis. He notes that a biopsy of the involved portion of the colon mucosa is pending for microscopic evidence of malignancy. Dr. A e-mailed Dr. B requesting that he call back to provide timely follow-up, which Dr. B did later that day.

Table 1 shows illustrative portions of the associated Communication Process Content Table.

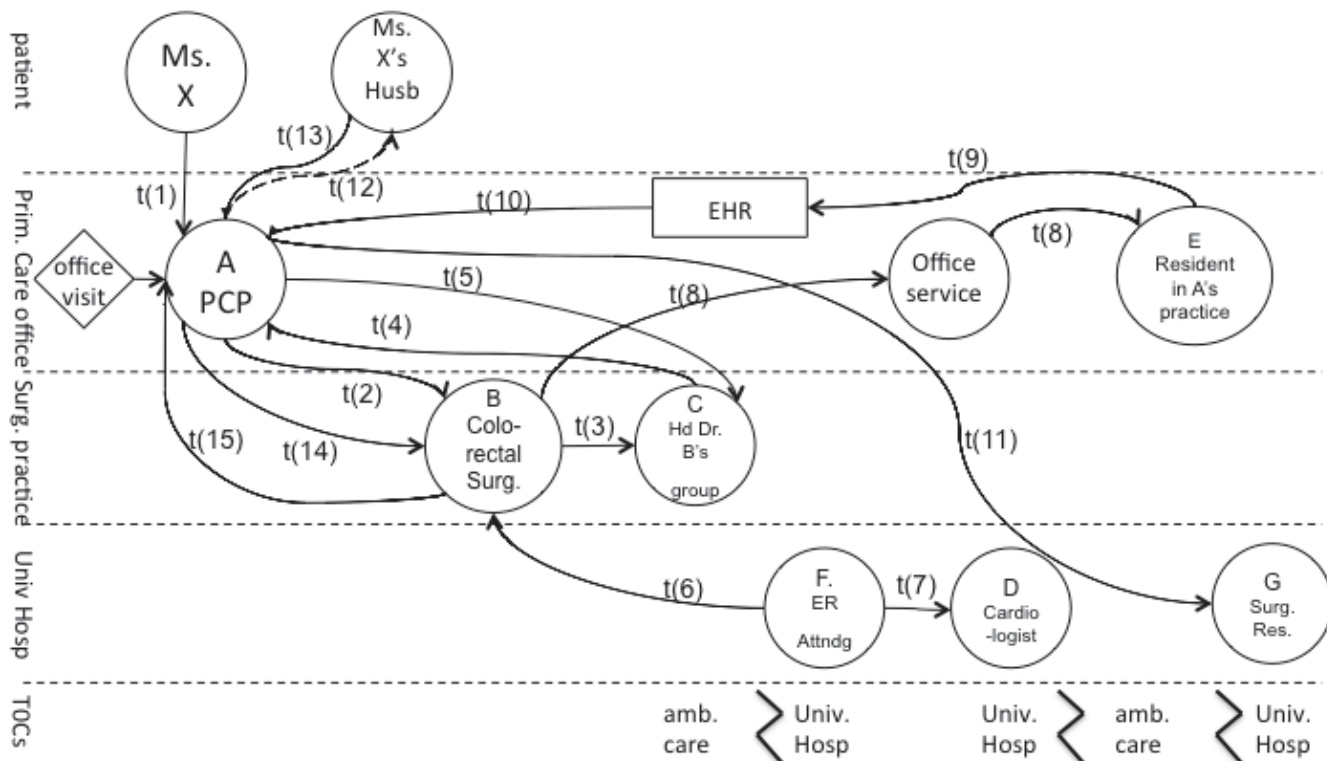


Figure 3. Coordination Process Diagram of Second Example Arc

Table 1. Portions of Communication Process Content Table for Figure 3

Sequence Number	1	2	3	...	6	7	8	9,10	...
Comm initiated from:	office encounter	O	O	...	Hospital	Hospital	Hospital	Hospital	...
Time rel to start (days.hrs)	0	1	2	...	15	15	18.1	18.6, 18.14	...
Who-Who	X->A	A->B	B->C	...	F->B	F->D	B->A's service->E	E->EHR->A	...
Intention	present problem	drill-down spec.	handoff	...	cross-elicite.	cross-elicite.	report back	report back	...
Modality	face-face	EHR	EHR	...	face-face	phone	Pager	EHR	...
PT gender, age, insurance	F,68,Medicare	F,68,Medicare	F,68,Medicare	...	F,68,Medicare	F,68,Medicare	F,68,Medicare	F,68,Medicare	...
PT Chronic Problems	Hypertention, Diabetes Mellitus, morbid obesity	Hypertention, Diabetes Mellitus, morbid obesity	Hypertention, Diabetes Mellitus, morbid obesity	...	Hypertention, Diabetes Mellitus, morbid obesity	Hypertention, Diabetes Mellitus, morbid obesity	Hypertention, Diabetes Mellitus, morbid obesity	Hypertention, Diabetes Mellitus, morbid obesity	...
PT Acute problems	intermittent abdominal pain	intermittent abdominal pain	intermittent abdominal pain	...	intermittent abdominal pain	intermittent abdominal pain	intermittent abdominal pain	intermittent abdominal pain, acute cholecystitis	...
PT medications				...	hyoscyamine	hyoscyamine	hyoscyamine	AB,hyoscyamine, intravenous fluid, pain medication	...
Studies/Pro. ordered	rectal exam	rectal exam	rectal exam, colonoscopy	...				CS	...
Studies/Proc. results	occult blood	occult blood	occult blood, results pending	...				blood cultures positive for gram negative rods	...
Communication content summary		A sends B an EHR task regarding pt's intermittent abdominal pain and + rectal exam	B sends C a task to detail pt's plan for colonoscopy	...	F consults B to evaluate pt. Sink admits pt to the hospital for observation	F consults D (cardiologist) for pt to have cardiac clearance while admitted in hospital	B sends page to E (in-patient covering resident for pt's PCP) detailing pt's admission.	EHR note: [full note text in narrative above, not here due to space limitations]	...



## 7 Conclusions

The Coordination Process Diagramming method introduced here is significant for several reasons:

1. It provides a method to break down, document, and visually represent arc of care coordination in a standardized and formal way that also provides a synoptic, visual, representation of an arc. This standardized representation can form the basis for common forms of analysis that can be used to identify:

- root causes of problems (such as failed communications, communications that should have occurred but were not attempted, and/or patient context information that was not propagated through the process as it should have been;
- structures or patterns of communication that differentiate arcs with good outcomes from those with less good or bad outcomes, to define measures, errors, root causes; and/or
- measures of care coordination that can be used to assess care coordination processes while they are on-going.

2. It can provide a foundation for simulation and model-based studies of care coordination processes. The structured decomposition provided by CPD can be used to design computer simulations of specific scenarios, or to generate and test new practices or policies for restructuring care coordination.

3. It can support the development of support systems to aid practitioners in care coordination. The CPDs can act as analogs to the use cases that are widely employed in software system design, providing well structured examples of real or hypothetical arcs that can drive the development of functionality of systems and tools that can help support, facilitate, and document care coordination processes.

## 8 Acknowledgement

The research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health, under award number 1R43AG044891. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## 9 References

- [1] Creswick, H., Westbrook, J. and Braithwaite J. (2009) Understanding communication networks in the emergency department. *BMC Health Services Research*, 9:247
- [2] Fineberg, H.V. (2012) A Successful and sustainable Health System—How to Get There from There. *New England Journal of Medicine* 366:1020-1027.
- [3] Gawande AA, Zinner MJ, Studdert DM, Brennan TA. (2003) Analysis of errors reported by surgeons at three teaching hospitals. *Surgery* 133:614-21.
- [4] Institute of Medicine (2000) *To Err Is Human: Building a Safer Health System*. Washington, DC: The National Academies Press.
- [5] Institute of Medicine (2012) *Living Well with Chronic Illness*. Washington, DC: The National Academies Press.
- [6] Press, M.J. (2014) Instant Replay — A Quarterback's View of Care Coordination. *N Engl J Med*; 371:489-491 August 7, 2014 DOI: 10.1056/NEJMp1406033
- [7] Scott, J., Tallia, A., Crosson, J.C., Orzano, J., Stroebel, J., DiCicco-Bloom, B., O'Malley, D., Shaw, E., Crabtree, B. (2005) Social Network Analysis as an Analytic Tool for Interaction Patterns in Primary Care Practices. *Annals of Family Medicine* 3(5):443-7
- [8] Stille, C., Jerant, A., Bell, D., Meltzer, D., and Elmore, J. (2005) Coordinating Care across Diseases, Settings, and Clinicians: A Key Role for the Generalist in Practice. *Annals of Internal Medicine* 14(9):700-708
- [9] Valente T.W. (2010). *Social Networks and Health: Models, Methods, and Applications*. NY, Oxford.
- [10] Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- [11] Weiner, S., Barnet, B., Cheng, T., and Daaleman, T. (2005) Processes for Effective Communication in Primary Care. *Annals of Internal Medicine* 142:709-714
- [12] Zachary, W., Maulitz, R., Iverson, E., Onyekwelu, C., Risler, Z., and Zenel, L. (2013). A Data Collection Framework for Care Coordination and Clinical Communications About Patients (CAPs). *Proceedings 2013 Symposium on Human Factors in Health Care: Advancing the Cause*. Santa Monica, CA: Human Factors and Ergonomics Society.
- [13] Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology*, 38(1), 188-230.

# Toward a Preliminary Layered Network Model Representation of Medical Coding in Clinical Decision Support Systems

H. Mallah<sup>1</sup>, I. Zaarour<sup>2</sup>, A. Kalakech<sup>2</sup>

<sup>1</sup> Doctoral School of Management Information Systems, Lebanese University, Beirut, Lebanon

<sup>2</sup> Management Information Systems Department, Lebanese University, Beirut, Lebanon

**Abstract** - *The aim of this research is to describe the impact of Medical Codes (MC) on building a successful Clinical Decision Support System (CDSS) that helps in structuring a beneficial Information Model, we highlight a preliminary Architectural Layered network model to assist health care givers and researchers in representing their work in a unified manner. Via a descriptive methodology, we incorporate the relationship between (MC) and (CDSS), we discussed a brief state of art of (CDSS) as well as medical coding as a process of converting descriptions of medical diagnoses and procedures into universal medical codes and numbers. Integrated with CDSS, medical codes are illustrated as an advanced search engine in the framework of Health Information System (HIS). We revealed the factual influence of health electronic information and proved the necessity of computerizing health care practices through the harmonization between the two categories of medical coding, medical classification (statistics) and nomenclatures (terminologies), and clinical decision support systems. The originality of our study is engaging medical coding and clinical decision support under cohesive scope, which was lacking for investigation. We relied on a plain case study to corroborate the ability of merging different types of medical coding in transforming health information from unstructured to richly structured data, to be used in building the medical knowledge base of the CDSS.*

**Keywords:** Medical Coding, Clinical Decision Support System, Layered Network Model, ICD, SNOMED CT, LOINC.

## 1 Introduction

Many technologies were invented and implemented in the human health domain in order to avoid falling in gaps since errors in such domain are extremely difficult to be corrected. The worldwide healthcare systems are facing many challenges, including soaring costs and failure in delivering the optimal patient care due to the

increase in medical errors, poor trainings to human resources and their derivatives. Many studies were done on how to computerize medical processes, especially critical decisions in medical fields, using Artificial Intelligence. From these needs, Clinical Decision Support System (CDSS) and Medical classification or medical coding gave birth after the development of the Electronic Health Record (EHR) and the Computer-based Physician Order Entry (CPOE). CDSS is the key of success of clinical information systems designed to assist health care providers in decision making during the process of care. This means that a clinician would cooperate with a CDSS to help determine diagnosis, analysis of patient data.

A significant breach exists between genuine clinical practice and optimal patient care. Several studies have shown that the quality of health care is variable and often insufficient <sup>[1, 2]</sup>. To deal with these deficiencies in providing care, healthcare institutions are gradually more depending on computer aided solutions such as clinical decision support systems (CDSS) that provide clinicians with patient-specific assessments for recommendations to aid clinical decision making <sup>[3,4]</sup>.

With the increased focus on the prevention of medical errors that has previously took place, computerized physician order entry (CPOE) systems, joined with CDSS, have been proposed as a key element of systems' approaches to improving patient safety<sup>[5,6]</sup>. If used appropriately, CDSS have the possibility to change the approach in which medicine has been taught and practiced.

The lack of standards in collecting and maintaining data has been a key barrier to electronic connectivity and transmission in healthcare <sup>[7]</sup>. Coordination between standard clinical terminologies and classifications represent a common medical language, allowing clinical data to be efficiently utilized. Therefore, standard clinical terminologies and classifications, with maps to link them, must be incorporated into EHR systems to achieve system interoperability and effective CDSS. Therefore, medical coding, closely tied in with the process of medical billing, is

an important feature to the health care industry and a well-structured knowledge base that contains accurate medical codes should be highly dependable to help CDSSs deliver appreciated outcomes.

This paper is a descriptive study that aims to illustrate the impact of Medical Codes (MC) on building a successful Clinical Decision Support System (CDSS) that helps in structuring a beneficial Information Model to assist health care providers and researchers in representing their work in a unified manner. The genuineness of our study is represented by fitting medical coding and clinical decision support system under cohesive scope, which was lacking for analysis. Illustrated by a case study, our objective is validating the ability of different medical coding types in transforming health information from unstructured to richly structured data. The assimilation done between medical codes from different categories leads to represent an integrated and hierarchal layered network representation consisting of five major layers.

In the first section, we will highlight the importance of Medical Coding. While in the second section, we will go over the integration of the Medical coding with Clinical Decision Support Systems to show its impact on having successful CDSS. A brief case study will be revealed in the third section, leading to the results and discussion in the fourth section.

## 2 Medical Coding: Classifications and Terminologies

**Medical coding** is the process of converting descriptions of medical diagnoses, clinical terms, tests observations, clinical findings and procedures into universal medical codes and numbers. The diagnoses and procedures are usually taken from a variety of sources within the patient health care record, such as the transcription of the physician's notes, laboratory results, radiologic results, and other sources. There are country specific standards and international classification systems.

Medical codes are mainly divided into 2 groups: Statistical Classifications and Nomenclatures or Terminologies <sup>[8]</sup>. A **statistical classification** brings together similar clinical concepts and groups them into categories (Mainly according to the medical specialty: Circulatory System, Respiratory System, Nervous system, etc...). An example of this category is the International Statistical Classification of Diseases and Related Health Problems (known as ICD). For example, the ICD 9 code of Tachycardia is 785.0. In a **nomenclature**, there is a separate listing and code for every clinical concept. So, in the previous example, each type and concept of the tachycardia would have its own code. This makes nomenclatures cumbersome for compiling health statistics.

Terminologies and classifications are designed for distinctly different purposes and satisfy different user data necessities. Classification systems such as ICD-9-CM, ICD-10-CM, and ICD-10-PCS group together similar diseases and procedures and organize correlated entities for effortless retrieval. They are typically used for external reporting requirements or other uses where data aggregation is beneficial, such as measuring the quality of care, monitoring resource utilization, or processing claims for reimbursement. Statistical classifications are considered “output” rather than “input” systems and are not proposed or designed for the major documentation of clinical care. Moreover, medical classifications are used for many other aspects including but not limited to conducting research, epidemiological studies, clinical trials, setting health policy, designing healthcare delivery systems, managing care and disease processes and tracking public health and risks. <sup>[9, 10, and 11]</sup>

On the other side, reference terminologies or nomenclatures such as Systematized Nomenclature of Medicine--Clinical Terms (SNOMED-CT) are “input” systems and codify the clinical information gathered in an EHR during the patient care encounter. This type of medical coding can be used for mainly getting access to complete and clear clinical data with links to medical knowledge for real-time clinical decision support. It also facilitates information exchange between providers thus fastening care delivery and minimizing replica testing and prescribing. And it can be used in retrieving information to generate expert alerts (e.g., allergy alerts, warnings of potential drug interactions or abnormal test results). In addition to that, medical terminologies grants access to standards of care for benchmarking, improving quality of care, measuring outcomes, developing and monitoring pay-for-performance programs, and measuring performance. <sup>[10, 12]</sup>

Together, standard clinical terminologies and statistical classifications represent a common medical language, allowing clinical data to be effectively utilized and shared between EHR systems. Therefore, standard clinical terminologies and classifications, with plans to link them, must be incorporated into Electronic Health Record EHR systems, that are main components of successful CDSSs, to achieve system interoperability and the benefits of a national health information infrastructure.

## 3 How Medical Codes employed in CDSS

“CDSSs are typically intended to integrate a medical knowledge base, patient data and an inference engine in order to generate case specific advice to health care practitioners leading to improve health related outcomes”. Some key questions to ask when considering CDSSs are “whose decisions are being supported, what information is

presented, when is it presented, and how is it presented to the user?"<sup>[13]</sup>. [Osheroff et al.] put forward what they call the **"five rights"** of Clinical Decision Support System which provides a relevant summary of what is needed for effective release: "CDSS should be designed to provide the right **information** to the right **person** in the right **format** through the right **channel** at the right **time**."<sup>[14]</sup>

CDSS are mostly composed of three main parts. These parts (figure 1) are the knowledge base, the inference or reasoning engine, and a mechanism or User Interface to communicate with the user.<sup>[15]</sup> The knowledge base consists of compiled information that is regularly, but not always, in the form of if-then rules. Other types of knowledge bases might include probabilistic relations of signs and symptoms with diagnoses, or known drug – drug or drug – food interactions. The second part of the CDSS is called the inference or reasoning engine, which contains the formulas for projecting the rules or associations in the knowledge base on the live patient data. Finally, there should be a mean of communication, which is defined by a way of getting the patient data into the system and getting the output of the system to the user who will make the actual decision. Output to the clinician may come in the form of a recommendation or alert at the time of order entry, or, if the alert was triggered after the initial order was entered, systems of email and wireless notification have been employed.<sup>[16, 17]</sup> Effective, well-implemented, well-used CDSS, along with quality assurance, feedback, and human management have been flourishing in reducing errors and adverse events, improving compliance with quality measures in many organizations and assisting clinicians at the point of care<sup>[18]</sup>.

Accordingly one of the most important challenges to have a successful CDSS is the right data format. CDSSs functions include searches for specific diagnostic statements and exclusive anatomical site acronym terms and/or abbreviations within any document from any medical record outcome. Such documents can consist of structured data determined by healthcare providers documenting care in an organization's EHR system(s) or formless data driven by healthcare providers documenting care in an organization's opposite EHR systems, such as text data generated by an organization's Dictation / Transcription / Speech Recognition (Voice / Text / Speech) systems. Such functions also include identifying both ICD-9-CM (the International Classification of Diseases, Ninth Revision, Clinical Modification) and/or ICD-10-CM/PCS (Procedure Coding Systems) obtainable on admission, principal, and secondary diagnoses and procedures for hospital inpatient documents, CPT-4 principal and secondary procedures for outpatient documents, and even nomenclature codes, such as SNOMED-CT, LOINC and RxNorm, for clinical, laboratory and pharmaceutical documents, respectively.<sup>[19]</sup> Linking the resulting, selected medical code(s) back to the source documentation supporting the code selection is important for providing full traceability. In addition, being able to have code suggestions occur at the point-of-care, as

documentation is entered into the record, is remarkable. For instance, there were thousands of medication items from the NDC (National Drug Code) directory added to the medication table and over 68,000 problem items (derived primarily from ICD-10-CM) added to the problem table. The network model of the working database structure (figure 2) coupled with a database model are constructed to simplify data query, each item in the medication data dictionary and each item in the problem data dictionary are connected by a common key attribute, an indication. In medical field, the National Cancer Institute (NCI) defined an indication as "a sign, symptom, or medical condition that leads to the recommendation of a clinical treatment, a laboratory test, or a treating procedure".

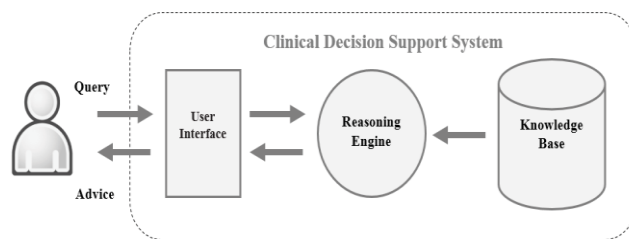


Figure 1: Architecture components of CDSS.

Each medication can be linked with its associated indications that can be represented as a group of relevant clinical problems.<sup>[20]</sup> Every single problem/diagnosis item in the problem table can be traced to a unique ICD-10-CM code and every single medication item in the medication table can be traced to a unique drug number RxN. Therefore, each ordered medication can be easily mapped by computer algorithm to one or more clinical problem(s) using recognized medication prescribing standards. This tracing methodology facilitates building and maintaining medical knowledge management (knowledge base in figure 1) and speeds up clinical decision support systems since it positively affects knowledge, decision support and information delivery axes that represent the main structural components of CDSSs.

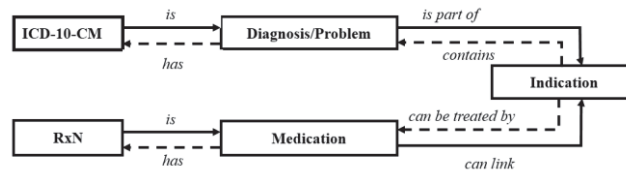


Figure 2: Sample network model

## 4 Case study

In this clinical case study, we will demonstrate how we can benefit from mixing between the two types of medical coding in enriching our medical knowledge base to be used in CDSS. We will make use two statistical classification coding methods that are ICD 10 and LOINC, in addition to SNOMED CT terminology. As per the International Health Terminology Standards Development Organization (IHTSDO), SNOMED CT owner and distributor, SNOMED CT wide-ranging coverage includes: clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimen. Similarly, the World Health Organization (WHO) enlarged ICD 10 to contain codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. Likewise, and since its initiation by the American Clinical Laboratory Association (ACLA) and the College of American Pathologists (CAP), LOINC database has prolonged to take account of not just medical and laboratory code names, but: nursing diagnosis, nursing interventions, outcomes classification, and patient care data set as well.

Let us take the following sample physician note:

*<< Mr. John Smith is a 27 years old male, with history of Dyslipidemia, came to the emergency department suffering of abdominal pain in the right lower quadrant. CT scan for abdomen and pelvis with IV contrast was performed. Patient was admitted to the hospital with Inflammatory Bowel Disease as admitting diagnosis. Colonoscopy with biopsy of colon was done showing ascending colon ulcer. After 24 hours, patient was discharged home, in a stable condition, on Lipanthyl 200mg one tab daily, according to Dyslipidemia medication review, and Flagyl 500 mgs two tabs daily for 5 days. >>*

From the above scenario, and in just using 3 medical coding categories SNOMED CT, ICD 10 and LOINC, we can get up to 59 different medical code derived from the medical terms written by the physician. The below tables of codes have been extracted using three main medical coding categories SNOMED CT (Figure 3.a), ICD 10 (Figure 3.b) and LOINC (Figure 3.c) for clinical terminologies, statistical classification and laboratory terms respectively.

SNOMED CT		
Clinical note	Code	Description
	371484003	Patient name
Mr. John Smith is a 27 years old male, with history of Dyslipidemia	102522009	Age 19 to 59 years (finding)
	884461000000102	Diagnostic history
	109041000119107	Complex dyslipidemia (disorder)
came to the emergency department suffering of abdominal pain in the right lower quadrant	4525004	Emergency department patient visit (procedure)
	163220003	On examination - abdominal pain
	301754002	Right lower quadrant pain (finding)
CT scan for abdomen and pelvis with IV contrast was performed	432370003	Computed tomography of abdomen and pelvis with contrast
	116154003	Patient (person)
Patient was admitted to the hospital with Inflammatory Bowel Disease as admitting diagnosis	8715000	Hospital admission (procedure)
	52870002	Admitting diagnosis
	52870003	Inflammatory bowel disease (disorder)
Colonoscopy with biopsy of colon was done showing ascending colon ulcer	446745002	Colonoscopy and biopsy of colon
	46040000	Ulceration of colon (disorder)
After 24 hours, patient was discharged home, in a stable condition	366471000000104	Reduction in length of hospital stay
	306689006	Discharge to home
	359746009	Patient's condition stable (finding)
on Lipanthyl 200mg one tab daily, according to Dyslipidemia medication review	418306005	Simplify home medication routine (procedure)
	409303000	Fenofibrate 200mg m/r tablet (product)
	473234001	Dyslipidemia medication review (procedure)
and Flagyl 500 mgs two tabs daily for 5 days	324523009	Metronidazole 500mg tablet (product)
	261774000	Duration of treatment (qualifier value)

Figure 3.a: SNOMED CT codes

ICD 10		
Clinical note	Code	Description
Mr. John Smith is a 27 years old male, with history of Dyslipidemia	E78.4	Other hyperlipidemia
came to the emergency department suffering of abdominal pain in the right lower quadrant	R10.2	Pelvic and perineal pain
	R10.30	Lower abdominal pain, unspecified
	R10.31	Right lower quadrant pain
CT scan for abdomen and pelvis with IV contrast was performed	N/A	N/A
Patient was admitted to the hospital with Inflammatory Bowel Disease as admitting diagnosis	K51.3	Ulcerative (chronic) rectosigmoiditis without complications
	K50.0	Crohn's disease of small intestine without complications
	K50.9	Crohn's disease, unspecified, without complications
	K51.2	Ulcerative (chronic) proctitis without complications
	K50.1	Crohn's disease of large intestine without complications
	K50.8	Crohn's disease of both small and large intestine without complications
	K52.1	Toxic gastroenteritis and colitis
	K52.0	Gastroenteritis and colitis due to radiation
	K51.8	Other ulcerative colitis without complications
	K51.0	Ulcerative (chronic) pancolitis without complications
	K52.9	Noninfective gastroenteritis and colitis, unspecified
	K52.81	Eosinophilic gastritis or gastroenteritis
	K51.4	Inflammatory polyps of colon without complications
	K52.2	Allergic and dietetic gastroenteritis and colitis
	K51.9	Ulcerative colitis, unspecified, without complications
K51.5	Left sided colitis without complications	
Colonoscopy with biopsy of colon was done showing ascending colon ulcer	K63.3	Ulcer of intestine
After 24 hours, patient was discharged home, in a stable condition	N/A	N/A
on Lipanthyl 200mg one tab daily, according to Dyslipidemia medication review	N/A	N/A
and Flagyl 500 mgs two tabs daily for 5 days	N/A	N/A

Figure 3.b: ICD 10 codes

LOINC			
Clinical note	Code	Description	
Mr. John Smith is a 27 years old male, with history of Dyslipidemia	N/A	N/A	
came to the emergency department suffering of abdominal pain in the right lower quadrant	N/A	N/A	
CT scan for abdomen and pelvis with IV contrast was performed	44115-4	Abdomen and Pelvis CT	
	42274-1	Abdomen and Pelvis CT W and WO contrast IV	
	36813-4	Abdomen and Pelvis CT W contrast IV	
	72250-4	Abdomen and Pelvis CT W contrast PO and W contrast IV	
Patient was admitted to the hospital with Inflammatory Bowel Disease as admitting diagnosis	N/A	N/A	
	18746-8	Colonoscopy study	
Colonoscopy with biopsy of colon was done showing ascending colon ulcer	28022-2	Colonoscopy Study observation Narrative	
	28023-0	Colonoscopy Study observation	
	28033-9	Colonoscopy, fluoro stoma Study observation Narrative	
	28034-7	Colonoscopy, fluoro stoma Study observation	
	60515-4	Rectum and Colon CT 3D W air contrast PR	
	67166-9	Sigmoidoscopy or colonoscopy [PhenX]	
	67172-7	Adenomas on sigmoidoscopy or colonoscopy [PhenX]	
	67220-4	In the past 2Y, have you had a virtual CT colonoscopy [PhenX]	
	67221-2	In the past 2Y have you had a colonoscopy [PhenX]	
	67223-8	Initial reasons you had a colonoscopy or sigmoidoscopy [PhenX]	
	72531-7	Rectum and Colon CT 3D W contrast IV and W air contrast PR	
	After 24 hours, patient was discharged home, in a stable condition	N/A	N/A
	on Lipanthyl 200mg one tab daily, according to Dyslipidemia medication review	N/A	N/A
	and Flagyl 500 mgs two tabs daily for 5 days	N/A	N/A

Figure 3.c: LOINC codes

## 5 Results and discussion

From the previously illustrated examples, we can conclude that as we increment the number of medical coding methods (from both medical coding classes), in the process to medical coding, as we will be able to extract more codes from unstructured data and subsequently structuring most of the formless data. In such a way, we will be able to enrich and enlarge the knowledge base used in clinical decision support systems.

As a result, the assimilation done between medical codes from different categories leads us to a depth and breadth network model (figure 4) which in terms represents an integrated and hierarchal layered network representation consisting of five major layers as follows: 1- Medical Coding Types, 2- Categories Coded, 3- Subcategories Coded, 4- Medical Terms Coded and 5- Protocols / Rules / Guidelines / Events ....The first layer represents the set of medical coding types from different categories to be merged (SNOMED CT, ICD 10, LOINC, etc...). Where the second layer stands for the set categories or classes to be coded using the coding types from the first layer (Findings, Body Structure, Diagnosis, Laboratory test, etc...). While the

subcategories of the previously listed set of categories (in layer number 2) will fall in the third layer (for Laboratory tests: Pathology, Hematology, Bacteriology, Chemistry, etc...). Nevertheless, the fourth layer will correspond to the coded medical terms of each subcategory listed in the preceding layer (for Chemistry Subcategory: CBC, Chem10, RBC, etc...). Finally, to end with the set of Protocols / Rules / Guidelines / Events... in the fifth layer (if SC1=x and a=Y then Run [1, 2 and 3]).

These outcomes (Protocols / Rules / Guidelines / Events) generated within the fifth layer in our model represent the core assets in any knowledge based Clinical Decision Support System. Therefore, the more we have coded terminologies from unstructured health related information the more we can resourcefully influence the knowledge, the decision support and the information delivery axes that stand for the key structural components of CDSSs. Where if we use coded information, building and managing medical knowledge base will be easier, the Reasoning Method will speed up and the CDSS response time will be faster at the point of care.

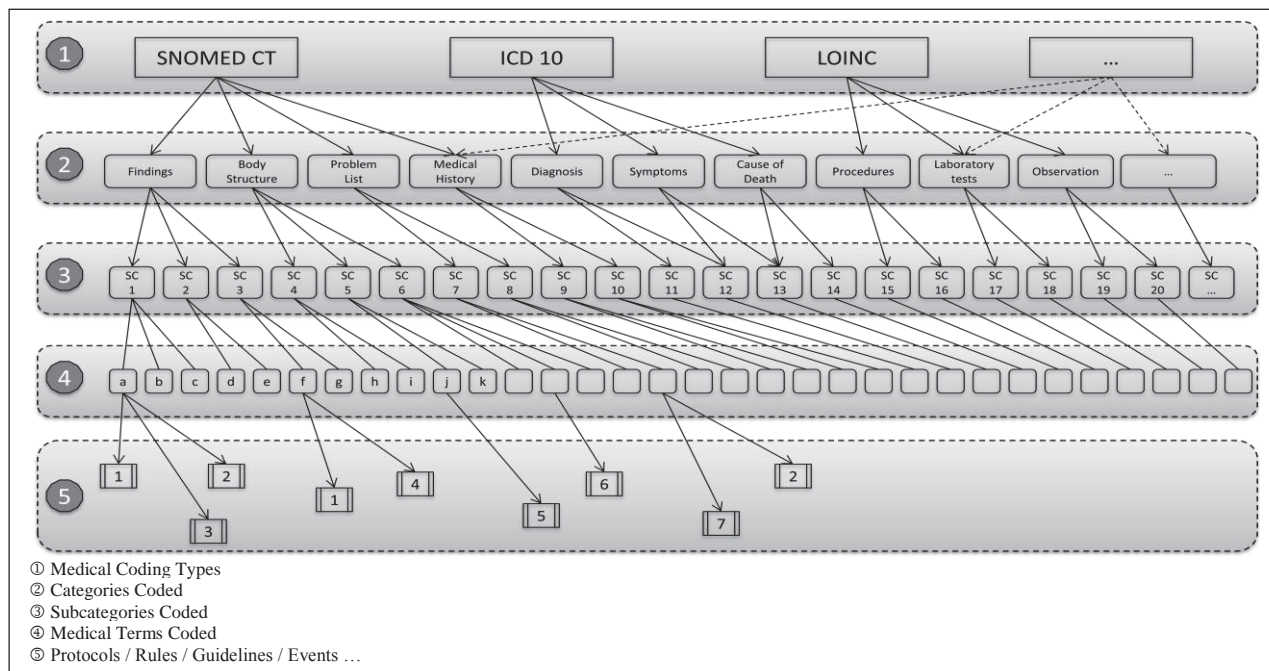


Figure 4: Depth and breadth network model

## 6 Conclusion

The inclusion of clinical terminologies into electronic health record systems is a significant step in the creation of information systems capable of monitoring quality and leading the practice of evidence-based medicine. A standard clinical terminology provides standardization of clinical terms, thus supporting easy transmission of patient data

across information systems. But the optimal value of the health information contained in an EHR system will only be grasped if both statistical classification and nomenclatures involved in the plan are up to date and accurately reflect the current practice of medicine. Together, standard clinical terminologies and classifications represent a common

medical language that allows clinical data to be shared between EHR systems. Therefore, standard clinical terminologies and classifications, with plans linking them, must be suited into EHR systems in order to attain system interoperability and fully benefit from CDSS, and that what was discussed in the first two sections of our study. In the third and fourth sections, we revealed a case study to back up the capability of amalgamating different types of medical coding in renovating health information from unstructured to abundantly structured data, to be used in building the medical knowledge base of the CDSS. In the last section and after the harmonized combination done between medical codes from different categories, we architected an integrated and hierarchal five layered network model. This model showed that the breadth and depth of both terminology and classification permit faster, reliable, and consistent retrieval of strong clinical information based on flexible queries in order to be used in enriching the medical knowledge base that symbolizes the core asset in any CDSS. Finally, this synchronization between medical coding (Statistical Classifications and Terminologies) and clinical decision support systems let healthcare institutions unchain the real power of health electronic information.

## 7 References

- [1] McGlynn EA, et al. (2003). The quality of health care delivered to adults in the United States. *The New England journal of medicine*, 2635-2645.
- [2] Jha AK, et al.; (2005). Care in U.S. hospitals-- the Hospital Quality Alliance program. *The New England journal of medicine*, 265-274.
- [3] Garg AX, et al.; (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA: The Journal of the American Medical Association*, 1223-1238.
- [4] Kawamoto K, et al.; (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ: British Medical Journal*, 765-765.
- [5] Kohn LT, et al.; (2000). To err is human: building a safer health system. *Journal of Vascular and Interventional Radiology*, P112-P113.
- [6] Bates DW et al.; (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medical errors. *JAMA: The Journal of the American Medical Association*, 1311-1316.
- [7] Wright, A., & Sittig, D.; (2008). A four-phase model of the evolution of clinical decision support architectures. *International Journal of Medical Informatics*, 641-649.
- [8] Madden R. et al., (2014). World Health Organization Family of International Classifications: definition, scope and purpose. *WHO, Technical Report*.
- [9] Margo I.; (2002). SNOMED Overview. *AHIMA's National Convention and Exhibit Proceedings*.
- [10] Benson T.; (2012). Principles of Health Interoperability HL7 and SNOMED (2<sup>nd</sup> Edition). *New York: Springer*.
- [11] Giannangelo K.; (2004). Clinical Vocabulary Basics. *AHIMA audio seminar*.
- [12] Greenberg M.; (2003). Recommendations for PMRI Terminology Standards. *NCVHS: The National Committee on Vital and Health Statistics, Technical Report*.
- [13] Berner E.S.; (2009). Clinical decision support systems: State of the Art. *AHQ: Agency for Healthcare Research and Quality, Technical Report (Publication No. 09-0069-EF)*.
- [14] Osheroff J.A.; (2009). Improving medication use and outcomes with clinical decision support: a step-by-step guide. *Chicago, IL: The Healthcare Information and Management Systems Society*.
- [15] Tan JKH, Sheps S.; (1998). Health decision support systems. *Gaithersburg: Aspen Publishers, Inc*.
- [16] Kuperman GJ, et al.; (1996). Detecting alerts, notifying the physician, and offering action items: a comprehensive alerting system. *Proceedings AMIA Annual Fall Symposium*, 704-708.
- [17] Shabot MM, et al.; (2000). Wireless clinical alerts for physiologic, laboratory and medication data. *Proceedings AMIA Symposium*, 789-793.
- [18] Berner, Eta S., ed.; (2007). Clinical Decision Support Systems. *New York, NY: Springer*.
- [19] Chiang S. Jaoe & Daniel B. Hier; (2010). Clinical Decision Support Systems: An Effective Pathway to Reduce Medical Errors and Improve Patient Safety. *Decision Support Systems* (pp. 121-138). *INTECH Open Access Publisher*.
- [20] Dineviski D. et al.; (2011). Clinical Decision Support Systems. *Telemedicine Techniques and Applications* (pp. 185-210). *INTECH Open Access Publisher*.

# Health Records on the Cloud: A Security Framework

Miloslava Plachkinova, Ala Alluhaidan, and Samir Chatterjee

Center for Information Systems and Technology, Claremont Graduate University, Claremont, CA, USA

**Abstract** - *The current study investigates the process of selecting a cloud service provider for implementing an electronic health record (EHR) system on the cloud from a security standpoint. This is an important issue because many eligible physicians still do not use EHRs and thus cannot meet the meaningful use criteria and receive incentives provided by the Centers for the Medicare and Medicaid Services (CMS).*

*To facilitate the process of selecting a cloud provider, we propose a framework focusing on the security issues related to implementing EHR systems on the cloud. The framework targets evaluators from healthcare practitioners and is designed to be a comprehensive tool for improving decision-making. The current study contributes to the literature as it utilizes design science methods to healthcare information systems and provides an overview of the current state of EHR, cloud computing, and security. The study also discusses many practical implications and recommendations for healthcare.*

**Keywords:** Healthcare, Security, Cloud Computing, Management, Decision Support, Design Science Research

## 1 Introduction

Modern healthcare has relied on technology as one of its pillars to success. Electronic health record (EHR) systems can improve the quality of care through clinical monitoring and by reducing rates of medical errors [1]. EHRs can also add value to medical practices by improving the clinical decision support systems [2]. Countless systems and applications exist to meet the needs of healthcare professionals and the vast amounts of data which they collect and store on the cloud [3]. This creates an enormous potential threat of undesired access to this information, Or the even more worrisome possibility of patients being intentionally harmed [4]. Cloud computing plays a significant role in the EHR implementation process because it improves the interoperability and access to information at minimal costs [5].

Security is an essential factor for every EHR system on the cloud [6]. Currently, the Centers for the Medicare and Medicaid Services (CMS) provide incentives for medical practitioners who are willing to participate in the process of implementing EHRs to achieve meaningful use. However, most eligible physicians still do not have EHR systems that demonstrate the meaningful use criteria [7] and thus cannot

take advantage of the offered incentives. One reason for this problem can be selecting a particular vendor or multiple vendors for implementing the EHR system. Currently, there are over 600 certified vendors [8] and choosing the right one presents a difficult task for many medical practices.

Even though selecting a vendor is a challenging issue, not much has been done in regards to addressing the medical practitioners' needs and to offering them practical advice on how to securely implement an EHR system onto the cloud. EHR implementation may come in two forms: (1) Vendors provide software and all the collected data resides within the medical facility; or (2) Vendors provide a web front-end but all processing and data is done remotely on the cloud. Prior literature review addresses in great detail cloud computing, healthcare information systems, and security as separate disciplines, but little research has been done on how to properly integrate them together. This lack of integration has led to significant difficulties when a decision has to be made in regards to the security of EHRs on the cloud. In spite of past research [9], there is still a great need for a practical sequence of steps to complete or check when selecting a suitable vendor that provides EHR services on the cloud.

This paper proposes a design science artifact, a security framework, which can be used by medical practices as they proceed on making a decision to implement an EHR system on the cloud. This framework summarizes the current state of research on cloud computing, healthcare IS, and security while providing healthcare professionals with a better process in selecting a cloud service provider (CSP).

Following main design science principles [10, 11], we aim to develop an online survey to evaluate the usefulness of the framework. The survey is planned to be distributed to healthcare providers. Overall, we consider the proposed artifact to be a useful tool for small-to-medium medical practices when selecting a CSP for health records on the cloud. A pilot study is also planned to provide constructive feedback to improve the effectiveness of the tool and to suggest barriers to implementing EHR systems on the cloud.

## 2 Related Work

Healthcare information systems have been extensively investigated by researchers [12-14]. Prior studies have clearly outlined the benefits of implementing such systems and their positive impact on the overall healthcare quality [15, 16]. These studies indicate that the potential of information



technology (IT) is far reaching in regards to how it can improve the existing hospital workflow and reduce medical errors and administrative efforts. It is difficult, especially for smaller medical practices, to keep pace with the latest trends in technologies. Financial limitations and a lack of resources also contribute to this trend. This is why finding a good process to select CSP can increase the success of such healthcare practices and can improve overall performance. For the purposes of this paper we use the classification of medical practices in terms of size (small, medium and large) as suggested by [17].

Clouds is defined as “large pools of virtualized resources which can be adjusted to a variable load scale and are exploited by a pay-per-use model.” [18] Thus, it allows for the incorporation of technologies into medical practices in an efficient manner [18]. According to the National Institute for Standards and Technology (NIST), there are five main characteristics of cloud computing: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service [19].

Private clouds can be used by a single practice where there is greatest control over the data and network. Public clouds are available to the general public but they provide no control. Hybrid clouds are a combination of both and they address most of the issues of each approach [20].

Also, there are three types of cloud services classifying the degree of control over the provided services. Software-as-a-Service (SaaS) offers no control to the client. Platform-as-a-Service (PaaS) grants control to subscribers. Infrastructure-as-a-Service (IaaS) offers strictly limited control to the deployed application and operating systems [21]. Clouds provides easy access to resources without requiring significant initial investments [22]. This is useful for small medical practices that do not have the capacity to build and support the required infrastructure. Yet, there is a lack of sufficient technical expertise and resources to choose the right CSP by Small-to-midsize practices.

Security of healthcare information systems requires specialized knowledge and experience, and practitioners in small-to-midsize practices may be unfamiliar with all the requirements and compliance issues. Research has been conducted in the past with regards to the security aspect of information systems [23]. However, the topic may be overwhelming for many medical professionals who lack the necessary expertise in the IT area. Healthcare security is also an aspect that researchers have focused on [24], and yet many questions remain unanswered. Some of them include: ensuring privacy and security of personal health records (PHR), overcoming barriers to adoption of PHR, and achieving meaningful use. Since many medical practices are emerging within the United States, securing patient information is one of the highest concerns to be addressed.

Healthcare cloud security has been investigated in the past without proposing viable solutions for it. Researchers have

identified healthcare security issues and applications of cloud computing [25], security and privacy concerns [26], and requirements [27]. This trend explicitly demonstrates the need for a solution to the secure implementation of EHR systems on the cloud.

### 3 Security Framework

The current paper proposes a security framework for health records on the cloud (Figure 1). This framework can be an efficient tool for small-to-medium healthcare for a better vendor selection process [28].

The framework is developed in accordance with design science principles suggested by [10, 11] to ensure its utility and usefulness. The framework we propose consists of five phases and represents a flow which can be successfully applied in the decision making process of selecting a cloud provider.

#### 3.1 Phase I: Requirements

Phase I consists of evaluating the technical capabilities of the CSP, and of financial and management analyses needed before committing to cloud services provider. These main principles have been identified in prior literature [29] to be of significant influence to the process of implementing EHR records on the cloud.

CSP needs to be carefully evaluated considering role-based services, so that only authorized personnel will have access to the necessary data and thus the problem of privilege abuse will be addressed [30]. The CSP also needs to provide secure user identification for monitoring and auditing access logs [31]. It is known that Cloud identity management is a challenge considering that data should be accessible easily regardless of location or device with a maintainable level of authenticity that is not obstructed by multiple methods of authentication. Centralized authentication with a single strong password is one way to address the previous challenge and provides a method for provisioning and de-provisioning [32]. Auditing logs can be used as evidence in an eventual investigation in cases of compromised data. Next, CSP's ability to provide sufficient capacity for the healthcare practice is of key importance. One of the main reasons for companies to use cloud services is the need for scalability. That saves them money and improves service costs in the long term. The CSP should be able to demonstrate capacity and data encryption. What mechanisms it uses to encrypt the data, how is the process being done, does it require public key infrastructure (PKI) or other encryption models? These are just some of the questions the medical practice needs to ask before committing to a contract with a specific CSP. In addition, securing the deletion of the data oftentimes, if not done properly, and can lead to data leakages or thefts. Monitoring system access should be done only by authorized individuals with verified credentials. The CSP needs to demonstrate clear backgrounds on all of its employees to reassure the integrity of its operations. Network security is

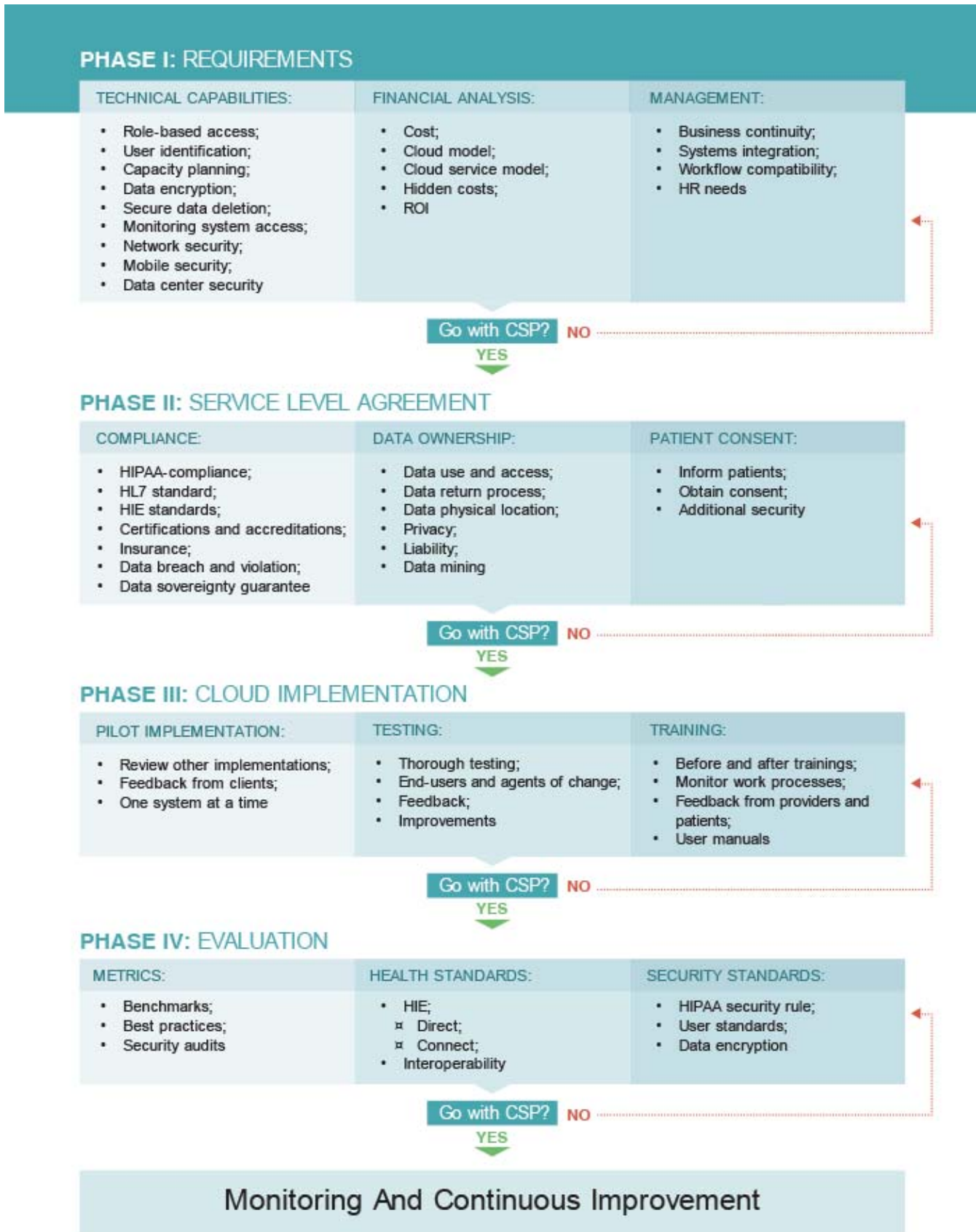


Figure 1. A Security Framework for Health Records on the Cloud

also important because it prevents attacks and intrusion from unauthorized third parties. The widespread use of smartphones and their growing capabilities [33] in modern healthcare requires a solid layer of mobile security [34]. And finally, the CSP needs to demonstrate the level of physical security of the data center where the healthcare data is stored.

The second aspect of Phase I is the need to complete a detailed financial analysis with respect to outsourcing data and/or services to a third party on the cloud. Discovering the needs of potential users (physicians, nurses, lab workers, clinical trial researchers, payroll, upper management, etc.) should be done as well. It is highly possible that the different groups have different needs and expectations. Yet, with many stakeholders involved in the process, it is possible to reach higher levels of user satisfaction, system quality and system use [35, 36]. Cost and Return on investment (ROI) needs to be considered because in some cases it may not be feasible to invest in an off-the-shelf EHR system or maybe the particular vendor may not be the most feasible option. Sometimes there are potential hidden costs that are not included in the initial estimation.

And finally, based on the collected information about the current technical capabilities, the elicited user requirements and the financial analysis, a business decision has to be made. Business continuity should be considered as part of the managerial evaluation process because data availability and interoperability between various healthcare providers are key factors for the CSP selection. Workflow compatibility guarantee seamlessly blend with previous structure without disrupting or changing the existing workflow processes at the healthcare practice. And the last management consideration should be related to the needs of the HR department that may involve access control for users group, the integrity, availability and confidentiality.

### 3.2 Phase II: Service Level Agreement (SLA)

Before committing to any CSP, the healthcare practice needs to discuss the eventual SLA and to make sure it has its best interests. With respect to compliance with legislation and best business practices, the CSP should be able to demonstrate Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) Act compliance, since it is essential for the receiving reimbursements according to the American Recovery and Reinvestment Act (ARRA). The usage of Health Level 7 (HL7) standards for messaging is also crucial, as it has become a de facto standard practice for the majority of US medical institutions. There are other Health Information Exchange (HIE) standards which are mandatory for all participating healthcare practices and the need to be reflected in the SLA. Additional certifications and accreditations are highly recommended as they ensure compliance with best practices and demonstrate the CSP's ability to meet industry standards. The CSP must also present insurance policies in case the data are breached, stolen or accessed by unauthorized parties. This will provide financial

guarantees to the healthcare provider and will increase the trust in the CSP. In case of a data breach or a violation, the SLA needs to address these problems and who will be responsible for them. The CSP should also guarantee the sovereignty of data, since the healthcare practice will not have physical access to it most of the time. Thus, it is important to ensure that data will not be compromised or any of their key functions (availability, integrity and confidentiality).

Additionally, data ownership needs to be included in the legal agreement. Ultimately, the customer should be the one to have control of how data is used and accessed and not the CSP. The healthcare practice needs to develop and enforce strict policies on this aspect and the CSP must comply with them. In cases when the customer does not want to use the services of the CSP, the legal contract needs to address the data return process – how long it will take, how it will be organized, who will have access to the data, etc. This is important to consider because the healthcare practice may not be satisfied with the cloud services provided and may want to terminate the contract. Since personal information about patients' healthcare will be stored on the cloud, it is important to also consider the physical location of the data center. It is recommended to be on US territory in order to avoid the possibility of national security issues related to data breaches, thefts or leakage. Some CSPs also use the data they store for data mining or selling it to other companies for marketing purposes. The legal contract must exclude all such actions as they violate the HIPAA and HITECH Act regulations.

The patients, whose data are going to be stored on the cloud, should be informed about this – how their data are being used, stored, and who will have access to them. Informing the patients and obtaining their written consent is of vital importance for the healthcare practice as this may lead to law suits, financial losses and lack of trust.

### 3.3 Phase III: Cloud Implementation

Phase III is concerned with the actual implementation of the EHR system on the cloud. First, a pilot implementation should be considered to avoid disrupting the business processes. This will provide an overview of the whole process and will demonstrate the CSP's ability to organize and support the process. Onsite visits can be helpful to obtain feedback directly from the CSP's customers and learn from their experience. After successfully completing the monitoring process and being satisfied with the results, the healthcare practice should consider implementing the cloud structure for one system at a time. That way the implementation process will proceed only if the previous system was successfully implemented.

A gradual implementation will allow thorough testing of each module and making sure that there are no bugs and the system functions as expected. The testing process can include stakeholders from various departments. Due to their specific needs, stakeholders may test a number of scenarios and look for specific problems. Integrated EHR systems usually

involve many individuals and processes, thus thorough testing plays an important role in the implementation process. Obtaining stakeholder feedback during the testing can be used not only to fix bugs, but also to suggest future improvements to the system's capabilities. Stakeholder involvement can also increase user buy-in and improve user satisfaction.

Training is the final step of the implementation process. It needs to be done with all individuals who have access to the data. Data will be shared with many other healthcare providers and it is important to maintain the integrity, authenticity and non-repudiation. Training the users to work with the new system will improve its acceptance and can also help with the proper use of the features and functionalities. Training users, before and after the implementation is complete, increases their skills and knowledge, and helps them to better utilize the capabilities of the system. Obtaining feedback not just from the medical personnel using the cloud solution, but also from the patients is necessary because the ultimate goal of the healthcare services on the cloud is to see improved quality of life for patients and make them more engaged in their own health. A detailed user manual should be provided by the CSP to the healthcare professionals so they can properly utilize the system and refer to it when necessary.

### 3.4 Phase IV: Evaluation

After implementing the CSP services or model, the next phase is evaluating whether the new system is useful and whether the CSP is delivering the services according to the SLA.

First, certain metrics need to be developed. They may be based on industry benchmarks or documented best practices. Such information is available in technical and medical journals and best practices are shared among the industry professionals. Also, regular security audits with specific criteria will be useful to make sure best practices are being followed.

The evaluation should consider how the system is implementing health standards. For example, the way the CSP deals with HIE in terms of Direct and Connect as means for communication in the National Health Information Network (NHIN); and interoperability between national systems and exchanging data according to the standards approved by the US government.

Evaluation of the existing health standards should be done regularly to reflect changes in legislation regarding application of CSP in healthcare. In addition, user standards regarding best practices in usability, usefulness and technical correction of the data need to be done during the evaluation process to adequately respond to any new improvements or developments in the healthcare cloud security field. And finally, evaluation of the data encryption model should be done because data need to be protected when transferred via the HIE. This is a must and a good way to protect the privacy of patients' data.

Evaluation from technical perspectives should also consider performance. Speed and capacity of the cloud as well as the frequency of backup are all requirements to maintain validity and reliability of the CSP. Privileges to access should be detailed. Secure multi-tenancy is essential where each customer see only their data and have no access to data belongs to others on the same cloud service [37]. CSP should be able to report logging investigations. Additionally, the portability of data is core criteria in making decision of any CSP to avoid Lock-in.

### 3.5 Phase V: Monitoring and Continuous Improvement

After implementing the CPS services or model, the final step for the medical practice is to perform continuous monitoring and make improvements to the system.

Making regular audits and searching for new and advanced practices regarding securing health information on the cloud are beneficial for the medical practice. These activities can increase the competitive advantage of the practice and create a sustained superior performance [38]. An assessment exercise can be performed during or after the implementation to identify the success of the project. Different techniques, individually or in combination, can be used to evaluate the implementation. Measuring results against goals, standards, and stated objectives will help managers to evaluate the performance of the project [39]. If the outcome of the evaluation is negative and the CSP fails to provide the necessary services, then looking for a new vendor is highly recommended.

## 4 Methodology

Our methodology is to use a qualitative survey and interviews to evaluate the proposed framework. The survey will be distributed to medical professionals. Face-to-face interviews with professionals will be conducted. Before that we plan to conduct a pilot test with graduate students and use their feedback to adjust the questions and improve the visual representation of the framework. Cognitive techniques will be used to code the main ideas provided by the respondents in the open-ended questions and interviews. Additionally, a frequency analysis to better understand the background of the respondents will be conducted. The online survey will comprise three sections. The first section informs the participants about the purpose of the study and obtains their consent. The second section contains background information in terms of the healthcare practice and the participants' level in it. In the third section, the framework is displayed and the participants will be asked to evaluate it based on a five-point Likert scale. Additionally, their feedback using the open-ended questions and interviews will be collected. For this section we used some of the survey questions developed by Motiwalla [40] as they were suitable for obtaining user feedback regarding the usefulness and usability of an artifact.

Since graduate students will complete the pilot survey, they need first to watch a short video online to learn more about

the security issues of EHR on the cloud, as we assumed they had no prior knowledge of the topic. This pilot study aims to make sure questions and explanations are clear enough and the security framework is easy to follow. The survey platform Qualtrics is the tool will be used to store and process the collected data.

Since this is an exploratory study, we plan to collect data from a convenience sample of healthcare professionals and information security experts who would be knowledgeable on the topic and can provide us with valuable feedback and ideas for improving the framework.

## 5 Discussion

Three possible limitations and recommendations for future research were identified in this study. First, this is an exploratory study on such a broad topic as securing health records on the cloud. We recommend follow-on action design studies to examine the effects of applying the security framework in a real healthcare environment. Specific metrics can be implemented to more accurately evaluate the utility of the proposed artifact. Second, utilizing qualitative techniques and using a case study approach may provide more details on the perspective of healthcare professionals regarding EHR systems on the cloud and the process of vendor selection. And third, the current study can be extended by considering cloud computing opportunities for mobile healthcare (mHealth). Security measures for smartphones are still under development and we recommend others to rely on a classification of knowledge such as the taxonomy of mHealth apps [41] in order to expand the scope of the current study.

## 6 Conclusion

The goal of this study is to propose a security framework for health records on the cloud. We expect, using the framework, to have a better scientific approach in selecting CSP. The study further demonstrates the effectiveness of having such a comprehensive approach to the decision-making process. Based on existing literature in healthcare IS, cloud computing and security, we developed a framework and described the methodology to evaluate the proposed artifact.

Our contribution to the existing literature is manifest in several aspects. First, we employ best practices in design science research and apply a design approach to solve an existing problem in healthcare IS. We follow recommendations by [10, 11] to demonstrate utility, usefulness, and the value of the proposed artifact. Second, we review the current state of healthcare information systems, cloud computing, and security to build the foundations of the framework. We summarize best practices and draw upon exemplar studies to provide a high-level concept of the CSP selection process. And third, we provide practical recommendations for improving healthcare IS and achieving the meaningful use criteria. Our artifact is based on prior scientific contributions and demonstrates high relevancy to a growing problem in healthcare IS.

The results of this study also have several implications for practice. By analyzing a medical practice's needs in terms of securely storing patient data on the cloud, this study provides an understanding of and an insight into the need of thoroughly investigating the CSP proposals in the context of information security. Evidence of the framework's application and usefulness highlight the importance and consequence of implementing such a tool in medical practices to improve their organizational processes. Healthcare professionals should focus on the security aspects to better protect health records on the cloud. By using the proposed framework, managers can make a more informed decision and select the vendor that can provide most capabilities at a reasonable price. That way, the medical practice can improve its performance, demonstrate meaningful use, and provide better quality of life for its patients.

## 7 References

- [1] Schiff, G.D., and Bates, D.W., "Can Electronic Clinical Documentation Help Prevent Diagnostic Errors?", *New England Journal of Medicine*, 362(12), 2010, pp. 1066-1069.
- [2] Romano, M.J., and Stafford, R.S., "Electronic Health Records and Clinical Decision Support Systems: Impact on National Ambulatory Care Quality", *Archives of Internal Medicine*, 171(10), 2011, pp. 897-903.
- [3] Lupse, O.S., Vida, M.M., and Stoicu-Tivadar, L., "Cloud Computing and Interoperability in Healthcare Information Systems", in (Editor, 'ed.'^eds.): *Book Cloud Computing and Interoperability in Healthcare Information Systems*, 2012, pp. 81-85.
- [4] Desai, A.M., and Mock, K., "Security in Cloud Computing", *Cloud Computing Service and Deployment Models: Layers and Management*, 208(2012),
- [5] Hammami, R., Bellaaj, H., and Kacem, A.H., "Interoperability for Medical Information Systems: An Overview", *Health and Technology*, 2014, pp. 1-12.
- [6] Rodrigues, J.J., De La Torre, I., Fernández, G., and López-Coronado, M., "Analysis of the Security and Privacy Requirements of Cloud-Based Electronic Health Records Systems", *Journal of medical Internet research*, 15(8), 2013,
- [7] Hsiao, C.-J., Decker, S.L., Hing, E., and Sisk, J.E., "Most Physicians Were Eligible for Federal Incentives in 2011, but Few Had Ehr Systems That Met Meaningful-Use Criteria", *Health Affairs*, 31(5), 2012, pp. 1100-1107.
- [8] <http://www.emrandhipaa.com/emr-and-hipaa/2012/04/11/over-600-ehr-vendors/>, accessed 01/10, 2014.
- [9] Lorenzi, N.M., Kouroubali, A., Detmer, D.E., and Bloomrosen, M., "How to Successfully Select and Implement Electronic Health Records (Ehr) in Small Ambulatory Practice Settings", *BMC Medical Informatics and Decision Making*, 9(1), 2009, pp. 15.
- [10] Hevner, A., March, S.T., Park, J., and Ram, S., "Design Science in Information Systems Research", *MIS Quarterly*, 28(1), 2004, pp. 75-105.
- [11] Hevner, A., and Chatterjee, S., *Design Research in Information Systems: Theory and Practice*, Springer, 2010.

- [12] Anderson, J.G., and Aydin, C.E., "Overview: Theoretical Perspectives and Methodologies for the Evaluation of Healthcare Information Systems": Evaluating the Organizational Impact of Healthcare Information Systems, Springer, 2005, pp. 5-29.
- [13] Berg, M., "Implementing Information Systems in Health Care Organizations: Myths and Challenges", International journal of medical informatics, 64(2), 2001, pp. 143-156.
- [14] Kushniruk, A., Borycki, E., Kuwata, S., and Kannry, J., "Predicting Changes in Workflow Resulting from Healthcare Information Systems: Ensuring the Safety of Healthcare", Healthcare Quarterly, 9(2006), pp. 114-118.
- [15] Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., and König, J., "Evaluation of Health Information Systems—Problems and Challenges", International journal of medical informatics, 71(2), 2003, pp. 125-135.
- [16] Poon, E.G., Jha, A.K., Christino, M., Honour, M.M., Fernandopulle, R., Middleton, B., Newhouse, J., Leape, L., Bates, D.W., and Blumenthal, D., "Assessing the Level of Healthcare Information Technology Adoption in the United States: A Snapshot", BMC Medical Informatics and Decision Making, 6(1), 2006, pp. 1.
- [17] Rittenhouse, D.R., Casalino, L.P., Shortell, S.M., McClellan, S.R., Gillies, R.R., Alexander, J.A., and Drum, M.L., "Small and Medium-Size Physician Practices Use Few Patient-Centered Medical Home Processes", Health Affairs, 30(8), 2011, pp. 1575-1584.
- [18] Vaquero, L.M., Rodero-Merino, L., Caceres, J., and Lindner, M., "A Break in the Clouds: Towards a Cloud Definition", ACM SIGCOMM Computer Communication Review, 39(1), 2008, pp. 50-55.
- [19] Mell, P., and Grance, T., "The Nist Definition of Cloud Computing (Draft)", NIST special publication, 800(145), 2011, pp. 7.
- [20] Zhang, R., and Liu, L., "Security Models and Requirements for Healthcare Application Clouds", in (Editor, 'ed.'^eds.): Book Security Models and Requirements for Healthcare Application Clouds, IEEE, 2010, pp. 268-275.
- [21] Odusote, B.O., and Ikhu-Omoregbe, N.A., "Towards a Well-Secured Electronic Health Record in the Health Cloud", Journal of Computing, 5(1), 2013,
- [22] Misra, S.C., and Mondal, A., "Identification of a Company's Suitability for the Adoption of Cloud Computing and Modelling Its Corresponding Return on Investment", Mathematical and Computer Modelling, 53(3), 2011, pp. 504-521.
- [23] Dhillon, G., and Backhouse, J., "Technical Opinion: Information System Security Management in the New Millennium", Communications of ACM, 43(7), 2000, pp. 125-128.
- [24] Appari, A., and Johnson, M.E., "Information Security and Privacy in Healthcare: Current State of Research", International Journal of Internet and Enterprise Management, 6(4), 2010, pp. 279-314.
- [25] Ahuja, S.P., Mani, S., and Zambrano, J., "A Survey of the State of Cloud Computing in Healthcare", Network and Communication Technologies, 1(2), 2012, pp. 12.
- [26] Löhr, H., Sadeghi, A.-R., and Winandy, M., "Securing the E-Health Cloud", in (Editor, 'ed.'^eds.): Book Securing the E-Health Cloud, ACM, 2010, pp. 220-229.
- [27] Gavrilo, G., and Trajkovic, V., "Security and Privacy Issues and Requirements for Healthcare Cloud Computing", in (Editor, 'ed.'^eds.): Book Security and Privacy Issues and Requirements for Healthcare Cloud Computing, 2012
- [28] Kushniruk, A., Beuscart-Zéphir, M.-C., Grzes, A., Borycki, E., Watbled, L., and Kannry, J., "Increasing the Safety of Healthcare Information Systems through Improved Procurement: Toward a Framework for Selection of Safe Healthcare Systems", Healthc Q, 13(1), 2010, pp. 53-58.
- [29] Kuo, A.M.-H., "Opportunities and Challenges of Cloud Computing to Improve Health Care Services", Journal of medical Internet research, 13(3), 2011,
- [30] Yu, S., Wang, C., Ren, K., and Lou, W., "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing", in (Editor, 'ed.'^eds.): Book Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing, IEEE, 2010, pp. 1-9.
- [31] Ruan, K., Carthy, J., Kechadi, T., and Crosbie, M., "Cloud Forensics": Advances in Digital Forensics VII, Springer, 2011, pp. 35-46.
- [32] <http://www.sans.org/reading-room/whitepapers/analyst/cloudy-chance-health-care-security-compliance-fundamentals-protecting-e-h-35055>
- [33] Boulos, M.N., Wheeler, S., Tavares, C., and Jones, R., "How Smartphones Are Changing the Face of Mobile and Participatory Healthcare: An Overview, with Example from Ecaalyx", Biomedical engineering online, 10(1), 2011, pp. 24.
- [34] Pharow, P., and Blobel, B., "Mobile Health Requires Mobile Security: Challenges, Solutions, and Standardization", Studies in health technology and informatics, 136(2008), pp. 697.
- [35] Gallivan, M.J., and Keil, M., "The User-Developer Communication Process: A Critical Case Study", Information Systems Journal, 13(1), 2003, pp. 37-68.
- [36] Hwang, M.I., and Thorn, R.G., "The Effect of User Engagement on System Success: A Meta-Analytical Integration of Research Findings", Information and Management, 35(4), 1999, pp. 229-236.
- [37] [http://www.cisco.com/c/dam/en/us/td/docs/solutions/Enterprise/Data\\_Center/Virtualization/Secure\\_Multi-Tenancy\\_Brief.pdf](http://www.cisco.com/c/dam/en/us/td/docs/solutions/Enterprise/Data_Center/Virtualization/Secure_Multi-Tenancy_Brief.pdf)
- [38] Porter, M.E., Competitive Advantage: Creating and Sustaining Superior Performance, Simon and Schuster, 2008.
- [39] Tulu, B., and Chatterjee, S., "A New Security Framework for Hipaa-Compliant Health Information Systems", in (Editor, 'ed.'^eds.): Book A New Security Framework for Hipaa-Compliant Health Information Systems, 2003, pp. 116.
- [40] Motiwalla, L.F., "Mobile Learning: A Framework and Evaluation", Computers & Education, 49(3), 2007, pp. 581-596.
- [41] Plachkinova, M., Andrés, S., and Chatterjee, S., "A Taxonomy of Mhealth Apps—Security and Privacy Concerns"

## **SESSION**

# **HEALTHCARE AND CLOUD INFRASTRUCTURE + MOBILE SYSTEMS + WEARABLES**

**Chair(s)**

**TBA**





# A Cloud-Based Infrastructure for Caloric Intake Estimation from Pre-Meal Videos and Post-Meal Plate Waste Pictures

Vladimir Kulyukin      Vikas Reddy Sudini  
Department of Computer Science  
Utah State University  
Logan, UT, USA  
[vladimir.kulyukin@ausu.edu](mailto:vladimir.kulyukin@ausu.edu)

Heidi Wengreen      Jennifer Day  
Department of Nutrition, Dietetics, & Food Sciences  
Utah State University  
Logan, UT, USA  
[heidi.wengreen@usu.edu](mailto:heidi.wengreen@usu.edu)

**Abstract**— Accurate caloric intake estimation is an open research problem in health informatics, dietetics, and nutrition management. Most research and development efforts to date have approached this problem by designing and developing automated vision-based methods to estimate caloric intake from static images. One reason why completely automated solutions underperform on some image sets is that they do not integrate nutritionists into the caloric intake estimation process. In this paper, a cloud-based infrastructure is presented that allows clients to submit pre-meal videos and post-meal plate waste pictures. Our approach recognizes the critical role of human nutritionists and keeps them integrated in the caloric estimation process.

**Keywords**—health data acquisition; caloric intake assessment; nutrition management; food intake; digital imaging; energy intake

## I. Introduction

Accurate caloric intake estimation is an open research problem in health informatics, dietetics, and nutrition management. Such chronic diseases as obesity, diabetes, and blockage of coronary arteries are related to mismanaged diets. Consequently, there is a growing need for robust methods for periodic caloric intake estimation. Most research and development efforts to date have approached this problem by designing and developing automated vision-based methods to estimate caloric intake from static pictures taken with mobile phones [1, 2]. Such methods attempt to automatically identify and, when possible, quantify consumed foods and beverages.

While the performance of automated approaches exhibits high percentages on selected food items [3], such approaches underperform on some image sets due to low image quality, inaccurate volume estimation, absence of reliable ground truth baselines, and highly variable food textures that poorly yield to automated classification.

One reason, relatively unexplored in the literature, why completely automated solutions may underperform on some data sets is that they do not integrate nutritionists into the caloric intake estimation process. Nutritionists are typically asked for feedback post-factum when the ground truth is needed to estimate the performance of a specific algorithm on a given set of images. Therefore, some nutritionists may feel disengaged, because they have little or no stake in the system.

Another reason why such systems may underperform is the limited ability of the target user to engage in continuous nutritional data collection and analysis on a regular basis, e.g., daily, weekly, or bi-weekly. The target users find it difficult to integrate nutritional data collection into their daily activities due to lack of time, motivation, or training, which causes them to turn off or ignore such digital stimuli as emails, phone calls, and SMS's.

To make nutritional data collection more manageable and enjoyable for the users, we have been developing a Persuasive NUTrition Management System (PNUTS) [4, 5]. PNUTS seeks to shift current research, training, and clinical practices in nutrition management toward persuasion, better integration of target users and nutritionists into dietary management, and community-oriented, context-sensitive nutrition decision support. PNUTS is inspired by the Fogg Behavior Model (FBM) [6], which states that motivation alone is insufficient to stimulate target behaviors. Even motivated users must have both the ability to execute behaviors and well-designed triggers to engage in that behavior at appropriate places or times. Toward this end, in this paper, we propose a cloud-based infrastructure that allows clients to submit pre-meal videos and post-meal plate waste pictures. Our approach recognizes the critical role of human nutritionists and keeps them completely integrated in the caloric estimation process. Automated image and video analysis approaches can be integrated into the infrastructure as needed.

Our paper is organized as follows. In Section II, we discuss related work. In Section III, we describe a cloud-based infrastructure for a PNUTS module that allows clients to submit pre-meal videos and post-meal plate waste pictures that are analyzed by human nutritionists for caloric intake. We describe the control flows in this module and discuss how the three main systemic roles (the client, the nutritionist, and the project coordinator) participate in these flows. In Section IV, we describe how the pilot version of this module has been deployed and integrated into *NDFS 4750: Transition to Professional Practice* taught at Utah State University (USU) in the spring 2015 semester. The class had thirty six students who acted as both client and nutritionist. Students did not have access to their own digital images and other digital images were accessed anonymously. The clients used the PNUTS

module to submit pre-meal videos and post-meal plate waste pictures. The nutritionists watched these data to estimate caloric intake. Two co-authors of this paper, Wengreen and Day, both Registered Dietitian Nutritionists acted as program coordinators to train nutritionists and to resolve inconsistencies in caloric intake estimations. In Section V, we analyze and discuss the results of the data collection and caloric intake estimation experiences with the system.

## II. Related Work

A variety of algorithms have been developed for vision-based caloric intake estimation. Bosch et al. [1] developed a system, called the mobile telephone food record (mpFR), to automatically identify and quantify consumed foods and beverages by analyzing images taken with mobile phones. The image analysis consists of image segmentation, food identification, and volume estimation. Pre- and post-meal images are used to estimate food and energy intakes. The reported recognition and energy intake estimation accuracy ranges from 50 to 90 percent.

Chen et al. [3] proposed a system for automated Chinese food identification and quantity estimation. The researchers use sparse coding in the SIFT and local binary pattern (LBP) feature descriptors combined with Gabor and color features to represent food items. A multi-label SVM classifier is trained for each feature. The trained classifiers are combined with the multi-class Adaboost algorithm [8]. The overall accuracy reported by the researchers is 68.3 percent. The experiments excluded transparent food ingredients such as pure water and cooked rice.

Kitamura et al. [7] proposed FoodLog, a web-based system that enables users to log their dietary intake by taking and uploading pictures of consumed foods and beverages. The system locates and analyzes the ingredients from the uploaded pictures and calculates the dietary intake according to a USDA food pyramid (usda.gov) that categorizes food into grains, vegetable, meat, beans, milk, and fruit. The researchers claim that FoodLog's performance is improved by personalized models created and updated dynamically via user feedback. The experiments showed that the accuracy of the food balance estimation was improved from 37 to 42 percent on average by personalized classifiers.

Hoashi et al. [9] proposed an automatic food image recognition system for 85 food categories by fusing various kinds of image features including bag-of-features (BoF), color histogram, Gabor features and gradient histogram with Multiple Kernel Learning (MKL). The researchers implemented a prototype system to recognize food images taken by mobile phone cameras. The MKL enabled the researchers to integrate various kinds of image features such as color, texture, and BoF representations. The researchers obtained an accuracy 62.5 percent classification rate for 85 food categories.

Yang et al. [10] proposed a new representation for food items based on pairwise statistics between local features computed over pixel-level segmentations of images into nine ingredient types: beef, chicken, pork, bread, vegetable, tomato, cheese, egg, and background. These statistics are collected in a multi-dimensional histogram, which is then used as a feature vector for a discriminative classifier. The Semantic Texton Forest (STF) [11] was used for image categorization and segmentation to generate soft labels for pixels based on

local such low-level characteristics as the colors of nearby pixels. The system was evaluated on the Pittsburgh Food Image Dataset (PFID) (<http://pfid.intel-research.net/>) and the proposed algorithm was compared with the two PFID baseline methods. The conducted experiments showed that the proposed method was significantly more accurate than the two baseline methods.

Martin et al. [2] proposed a method called the Remote Food Photography Method (RFPM). The method consists of camera-enabled cell phones with data transfer capability. Users take and transmit photographs of food selection and plate waste to researchers or clinicians for subsequent analysis. The RFPM allows clients to receive fast feedback about their energy intake from professionals at remote locations without going to a clinic. The RFPM was tested in controlled laboratory and free-living conditions, which allowed the researchers to do direct comparisons of the accuracy between laboratory and free-living conditions. The variety of food was limited and not necessarily representative of the participants' habitual daily intake.

Wang et al. [12] developed a new dietary instrument for assessing an individual's food intake with a hand-held personal digital assistant (PDA) with a camera and a mobile telephone card. The researchers applied a cross-sectional study design in a study of twenty-eight participants who were asked to keep 1-day weighed food records. Digital images of all recorded foods were obtained simultaneously and sent to registered dietitians via PDAs. The participants' opinions about the PDA method and two other methods were determined using a questionnaire. No significant differences were found between the PDA method when compared with the other two food recording methods for most nutrients. The survey showed that 57 percent of the participants indicated that the PDA method was the least burdensome of the three methods and the least time consuming to record daily diet.

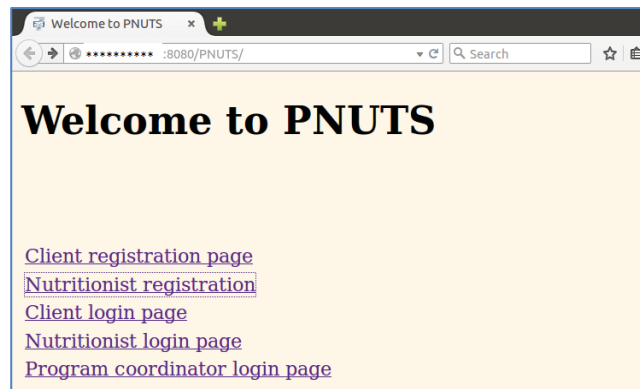


Figure 1. Home page of PNUTS

## III. Caloric Intake Estimation Infrastructure

In this section, we describe a cloud-based infrastructure module for PNUTS that allows clients to submit pre-meal videos and post-meal plate waste pictures. Our approach is similar to the approaches outlined in Martin et al. [2] and Wang et al. [12] in that it recognizes the critical role of human nutritionists and the limitations of completely automated food image analysis approaches. Unlike in these two systems, the data input in our infrastructure is not

confined to static images taken on mobile phones: the clients can submit not just static photos but also videos. Figure 1 shows the home page of the PNUITS module described in this paper and used in the pilot study. As can be seen from Figure 1, the PNUITS module has three main systemic roles: the client, the nutritionist, and the program coordinator. Each role is briefly described below.

A. Client

The client is any user of the system interested in obtaining accurate caloric intake estimations of consumed foods and beverages. The client uses a web-enabled device to register with the system. The client receives text reminders before the three major meals (breakfast, lunch, dinner) and has an option to sign up for additional reminders or tips. At breakfast, lunch, and dinner the client takes a short per-meal video and does an optional voice recording the foods and beverages to be consumed.

At the end of the meal the client takes a still picture of the plate waste. The client is required to have a standard plastic card object in all videos and pictures for subsequent caloric intake estimations. Figure 2 shows a snapshot from a pre-meal video with the sensitive information eliminated from the client's USU card. Figure 3 shows the corresponding post-meal plate waste picture. The pre-meal video and post-meal picture are uploaded by the client on a web page shown in Figure 4. These data together with the client id and a time stamp constitute an energy intake estimation request (EIER).



Figure 2. Snapshot from a Pre-Meal Video



Figure 3. Post-Meal Plate Waste Picture

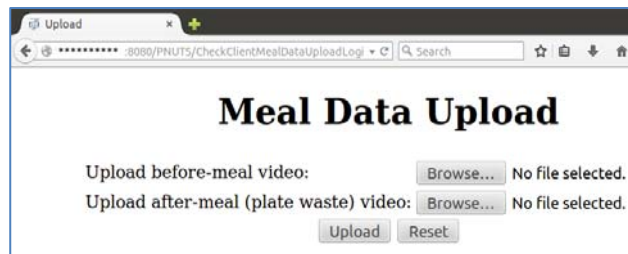


Figure 4. Client Meal Data Upload

B. Nutritionist

The nutritionist also registers with the system to review and evaluate EIERS. Figure 5 shows a web page where a nutritionist selects EIERS for evaluation. The nutritionist watches the videos and looks at the plate waste pictures. In computing the caloric intake the nutritionists use USDA's National Nutrient Database (ndb.nal.usda.gov/ndb). Nutritionists selected the appropriate food and estimated the amount eaten. Total kcalories consumed was computed by the NDB program and entered into the PNUITS module by the nutritionist. Figure 6 shows two caloric estimations of the meal data shown in Figures 2 and 3.

Request ID	Client ID	Before Meal	time	After Meal	time	Evaluation
52	a0184	data	2015-02-26 15:17:03.0	data	2015-02-26 15:17:03.0	evaluate
53	a0184	data	2015-02-26 15:23:00.0	data	2015-02-26 15:23:00.0	evaluate
66	a3164	data	2015-02-26 21:13:48.0	data	2015-02-26 21:13:48.0	evaluate
69	a8054	data	2015-02-26 21:20:29.0	data	2015-02-26 21:20:29.0	evaluate
72	A0149	data	2015-02-26 21:24:31.0	data	2015-02-26 21:24:31.0	evaluate
73	a7135	data	2015-02-26 21:24:33.0	data	2015-02-26 21:24:33.0	evaluate
74	a7135	data	2015-02-26 21:25:13.0	data	2015-02-26 21:25:13.0	evaluate
75	a8054	data	2015-02-26 21:25:37.0	data	2015-02-26 21:25:37.0	evaluate

Figure 5. Pending EIERS for Nutritionists

- Nutritionist ID: A\*\*\*\*\*2540
- Email ID: \*\*\*\*\*@gmail.com
- Estimation: [54 A01772540.txt](#)

Summary:

08539 Kashi Wheat Cereal 1 cup 336  
 01175 1% milk 1 cup 102  
 09433 Clementines 2 fruits 70

- Nutritionist ID: A\*\*\*\*\*9151
- Email ID: \*\*\*\*\*@gmail.com
- Estimation: [54 A01049151.txt](#)

Summary:

08539 Cereals ready-to-eat, Kashi Organic Promise, Cinnamon Harvest 1 serving 185  
 01082 milk, lowfat, fluid, 1% milkfat, with added nonfat milk solids, vitamin A and D 1 cup 102  
 09433 Clementines, Raw 2 each 70

Figure 6. Two Conflicting Estimations

An EIERS can be in one of the five states: *unprocessed*, *pending*, *processed*, *conflicting*, and *resolved*. A request is

*unprocessed* when it has been received by the system but has not been evaluated by any nutritionist registered with the system. A request is *pending* when it has been evaluated by one nutritionist. A *processed* request is a request that has been evaluated by two nutritionists without a conflict, i.e., their estimates agree with each other within ten percent of total calories at the eating occasions level, or by a program coordinator, another nutritionist with the authority to resolve conflicting caloric intake estimations. A *conflicting* request is a request that has been evaluated by three nutritionists whose evaluations disagree by more than 10 percent. A *resolved* request is a conflicting request that has been evaluated by a program coordinator. Resolved requests can be viewed by nutritionists but not re-evaluated.

C. Program Coordinator

Figures 7 and 8 show how the energy intake estimations (EIE) are made by nutritionists and, if necessary, resolved by program coordinators. When a nutritionist submits his or her estimation, the new EIE is matched against the list of EIEs for that request. If there are no existing EIEs for that request, the system does nothing and saves the EIE in the EIE database.

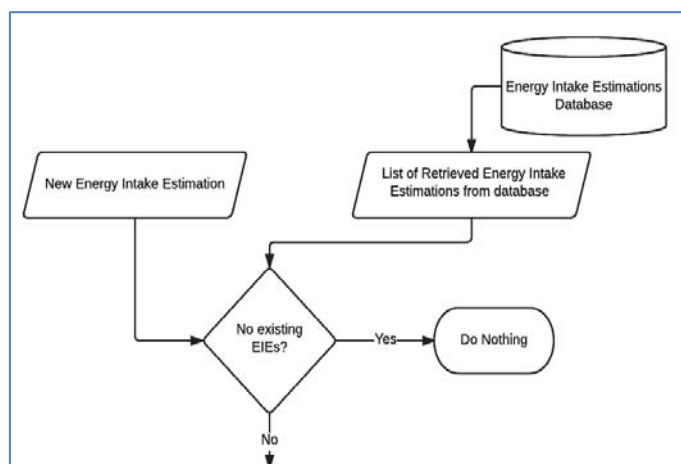


Figure 7. No Existing EIEs for a given EIER

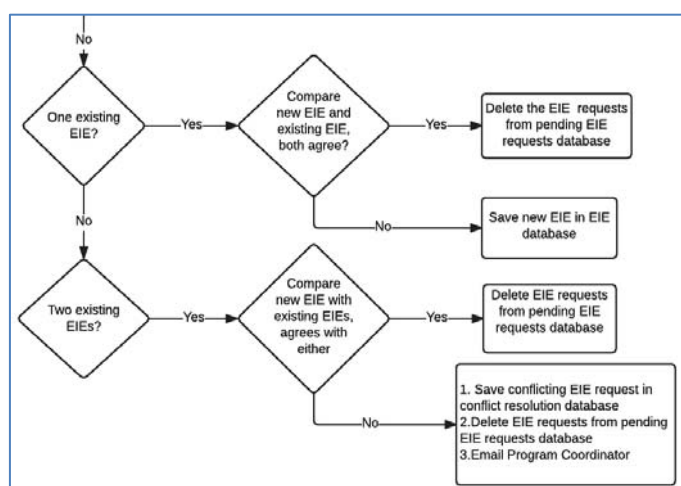


Figure 8. Matching new EIE against existing EIEs

If there is at least one existing EIE for that EIER, the control flow continues as shown in Figure 8. If there is only one existing EIE, the new EIE is compared with it. The comparison between two requests is computed as the relative difference between two numbers, i.e.,  $D = \frac{|x-y|}{\max\{x,y\}} 100$ , where x and y are two total caloric counts in the two EIEs being compared. For example, consider two EIEs given in Figure 6. The first EIE (EIE1) is submitted by the nutritionist whose ID is A\*\*\*\*\*2540. The second EIE (EIE2) is submitted by the nutritionist whose ID is A\*\*\*\*\*9151. The total caloric count of EIE1 is 336+102 +7 =508. The total caloric count of EIE2 is 185+102+70=357. The relative difference is  $\frac{|508-357|}{508} 100 = 29.72$ . Two EIEs are considered similar if their relative difference is at most 10 percent. Thus, the two EIEs given in Figure 6 are considered conflicting.

If the new EIE and the only existing EIE are not conflicting, the corresponding EIER is removed from the database of pending EIERs and considered resolved. If the new EIE and the only existing EIE are found to be in conflict, both are saved in the database of the existing EIEs for a given EIER.

If there are two existing EIEs for a specific request, then the new EIE is matched against both to see if there is agreement with either one. If there is an agreement, all EIEs are removed and the corresponding EIER is considered resolved. If there is no agreement, all three EIEs are deleted from the database of the existing EIEs, the corresponding EIER is deleted from the database of the pending EIERs, and all four records (three EIEs and one EIER) are saved in a conflict database. A program coordinator is notified about the conflict via email.

Request ID	Client ID	Before Meal	Time	After Meal	Time	Status
50	a0184	data	2015-02-26 15:05:17.0	data	2015-02-26 15:05:17.0	not resolved
51	a0184	data	2015-02-26 15:15:44.0	data	2015-02-26 15:15:44.0	not resolved
54	a4724	data	2015-02-26 20:42:17.0	data	2015-02-26 20:42:17.0	not resolved
56	a4724	data	2015-02-26 20:47:42.0	data	2015-02-26 20:47:42.0	not resolved
57	a3164	data	2015-02-26 21:00:14.0	data	2015-02-26 21:00:14.0	not resolved
58	a4724	data	2015-02-26 21:06:25.0	data	2015-02-26 21:06:25.0	not resolved
60	a3159	data	2015-02-26 21:09:05.0	data	2015-02-26 21:09:05.0	not resolved
62	a4724	data	2015-02-26 21:09:44.0	data	2015-02-26 21:09:44.0	not resolved
63	a4724	data	2015-02-26 21:10:35.0	data	2015-02-26 21:10:35.0	not resolved
64	a0104	data	2015-02-26 21:11:32.0	data	2015-02-26 21:11:32.0	not resolved
65	a0104	data	2015-02-26 21:13:23.0	data	2015-02-26 21:13:23.0	not resolved

Figure 9. List of Conflicting EIERs

The program coordinator is also a registered Dietitian Nutritionist who has the authority to resolve EIER conflicts. The system can have multiple program coordinators, each supervising a number of nutritionists. In the pilot version of the system described in this paper, there are two program coordinators.

Whenever a conflicting EIER request has been detected by the system, the program coordinator receives an email notification. When the program coordinator logs in, the program coordinator sees a list of conflicting EIERs, as shown in Figure 9. The program coordinator watches the same video data as the nutritionists who provided the EIEs and evaluates

the request. After the program coordinator evaluates the conflicting EIER, the EIER is considered resolved and is removed from the database of the conflicting requests.

D. Impelementation Details

We implemented the described infrastructure in Java using the Java servlets, Java Server Pages (JSPs), and JBoss (<http://www.jboss.org>). Our current cluster has two nodes: one master and one slave. All our databases are implemented with MySQL (<https://www.mysql.com/>). The JBoss Application Server (JBoss AS) is a free open-source Java EE-based application server. In addition to providing a full implementation of a Java application server, it also implements the Java EE part of Java. The JBoss AS is maintained by jboss.org, a community that provides free support for the server. JBoss is licensed under the GNU Lesser General Public License (LGPL).

IV. Pilot Deployment

The pilot version of the PNUITS module described in the previous section has been deployed and integrated into NDFS 4750: Transition to Professional Practice taught at Utah State University (USU) in the spring 2015 semester. 407 pre-meal videos, 407 post-meal pictures, and 84 caloric intake estimations.

Food Item	Estimated amount	Food Code (USDA)	Calories
orange	1 medium		0 - did not eat
Chobani FF pomegranate yogurt	1 serving cup	(Chobani website)	120
red onion	2 slices	11282	30
orange bell pepper	1/2 cup	11961	14
eggs	2	01123	143
feta cheese	1oz	01019	76
chopped parsley	1 Tbsp	11297	1
			Total Calories: 383

Figure 10. Sample Caloric Intake Estimation Form

All students enrolled in NDFS 4750 received caloric estimation training prior to acting as clients and nutritionists. Each student watched and evaluated fourteen training pre-meal and post-meal video pairs and was required to complete by hand a caloric intake estimation form shown in Figure 10. The ground truth for each video was provided by Jennifer Day. Each student's evaluation was compared with the ground truth and the students who differed from the ground truth by more than ten percent had to undergo additional training. The range of the ratio of EIEs to ground truth was 0.85 - 1.15 (mean = 0.98, standard deviation = .067). Of the thirty six students, thirty two (89%) provided EIEs within 10% of the ground truth. Four students provided EIEs that were not within 10% of the ground truth. One student provided mean EIEs that underestimated intake, and three students overestimated energy intake. After additional training all four students provided estimates within 10% of ground truth using digital data from five new eating occasions.

The client group used the PNUITS module to submit pre-meal videos and post-meal plate waste pictures. We have collected sixty one pre-meal videos, sixty one post-meal plate waste pictures, and one hundred twenty two EIEs. The nutritionists used the data to estimate caloric intake of each eating occasion. Jennifer Day, acted as program coordinator to resolve inconsistencies in caloric intake estimations. In addition, each client completed a paper-pencil three day food record of all foods eaten and submitted to the PNUITS program as pre- and post-meal digital data. Figure 11 shows part of a filled 3-day food record form filled by one of the student clients. We plan to use these data forms as the ground truth in our future experiments on automated food item identification from client videos.

Twenty eight students provided EIEs for from digital data of 61 eating occasions provided by this same set of students. The mean EIEs was 383 kcalories (range: 35 - 936). Agreement among the twenty eight students who provided EIEs from the digital data was examined with intra-class correlation coefficients (ICCs). Agreement was high among the twenty eight student nutritionists (ICC = 0.89, p<0.001). Agreement was lower for digital data from eating occasions that included a single food or five or more food items (ICC = 0.78, 0.77, respectively; p-value = 0.023, and 0.002, respectively). Agreement was highest for digital data from eating occasions that included two to four food items (ICC = 0.95, p<0.001). These values are at least as high as the ICCs reported when digital photography method is used in cafeteria settings and free living adults [13, 14].

Food Item	Description of Food/Beverage (Brand name, cooking method used, type i.e. low fat or low carbohydrate, if fortified, and any other additional information)	Amount	Unit
Cereal	Honey Bunches of Oats w/ Almonds	1	C
Milk	1%	1	C
Juice	Med - clementine	3	oz
Tea	herbal	1	cup
Sugar	Reg White	1	tea
Water	Bottled 20oz	10	oz
yogurt	yooplait - strawberry	1	loozy
Banana	Med size	1	ea
Water	Bottled		
Yogurt	plain, reg fat	1/2	C
eggplant	roasted in oven	3/4	C
Beef	90/10 ground	< 1/4	C
Parsley		< 1/4	C
onions	put in as onion rings green onions	< 1/8	C
garlic		< 1	TB

Figure 11. Part of a 3-Day Paper Record

## V. Discussion

The pilot version of the PNUTS infrastructure functioned flawlessly and enabled both the clients and the nutritionists to upload and analyze the meal data. There were two main complaints by the students who used the system as clients and nutritionists. The clients complained that the system did not allow them to submit partial caloric estimations with an option to complete them later. In the current implementation, the client can submit only complete estimations of each meal. The main complaint of the nutritionists was the necessity to use the USDA National Nutrient Database at [ndb.nal.usda.gov](http://ndb.nal.usda.gov). The students did not complain about the Nutrient Database program itself but about the need to go to a different web site, look for ingredients, enter in the amount eaten, then retype all the data into another web page.

We plan to address the first complaint by allowing the clients to submit partial caloric intake estimations. One challenge that we foresee is what to do with partial estimations when their requests have been completed by other nutritionists. For example, a nutritionist who comes back to complete an estimation may discover that two other nutritionists have submitted complete non-conflicting estimations of the same EIER. The second complaint will be harder to address as we do not our own internal database of all possible ingredients that may be found in submitted digital data.

In an informal qualitative survey conducted with the students they said that the system made them feel as part of a community of users and that they would use the system as registered nutritionists after they graduate. When asked about the caloric data entry difficulties, the students said these difficulties are compensated by not having to meet with the clients in person. The system has turned out to be a valuable training tool for undergraduate nutrition students and will likely be used in other USU NDFS classes in the fall 2015 semester.

Another direction that we plan to pursue in the future is automated identification of food items in pre-meal videos. Such identification will likely make data entry easier for nutritionists in that the nutritionists will not have to enter the names of food items manually.

## Acknowledgments

We are grateful to Dr. Sheryl Aguilar for letting us to pilot the system in her NDFS 4750 class that she taught at USU in the spring 2015 semester. We are also grateful to all students enrolled in NDFS 4750 who tested the system as clients and nutritionists for their effort and feedback.

## References

- [1] Bosch M., Zhu F., Khanna N., Boushey C.J., and Delp E. "Combining global and local features for food identification in dietary assessment." *IEEE transactions on Image Processing*. 2011:1789-1792. doi:10.1109/ICIP.2011.6115809.
- [2] Martin, C.K., Han, H., Coulon, S.M., Allen, H.R., Champagne, C.M., and Anton, S.D. (2009). "A novel method to remotely measure food intake of free-living people in real-time: The Remote Food Photography Method (RFPM)." *British Journal of Nutrition*, 101, 446-456. PMID: PMC2626133.
- [3] Chen, M.Y., Yang, Y., Chia-Ju Ho, C., Wang, S., Liu, S., Chang, E., Yeh, C., & Ouhyoung, M. "Automatic Chinese food identification and quantity estimation." In *Proceedings of SIGGRAPH Asia Technical Briefs (SA '12)*. ACM, New York, NY, USA, , Article 29 , 4 pages, 2012. DOI=10.1145/2407746.2407775.
- [4] Kulyukin, V., Zaman, T., and Andhavarapu, S. "Effective use of nutrition labels on smartphones." In *Proceedings of the 15th International Conference on Internet Computing and Big Data (ICOMP 2014)*, pp. 93 - 99, July 21-24, 2014, Las Vegas, NV, USA, CSREA Press, ISBN: 1-60132-227-1.
- [5] Kulyukin, V. and Zaman T. (2013). "Vision-based localization of skewed UPC barcodes on smartphones." In *Proceedings of the International Conference on Image Processing, Computer Vision, & Pattern Recognition (ICCV 2013)*, pp. 344-350, pp. 314-320, ISBN 1-60132-252-6, CSREA Press, Las Vegas, NV, USA.
- [6] Fogg, B.J. "A behavior model for persuasive design." In *Proceedings of the 4th International Conference on Persuasive Technology*. Article 40. ACM, New York, USA, 2009.
- [7] Kitamura, K., Yamasaki, T., & Aizawa, K. 2009. "FoodLog: capture, analysis and retrieval of personal food images via web." In *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities (CEA '09)*. ACM, New York, NY, USA, pp. 23-30. DOI=10.1145/1630995.1631001.
- [8] Freund, Y. & Schapire, R. "A Decision-theoretic generalization of on-line learning and application to boosting." *Journal of Computer Systems and System Sciences*, vol. 55, issue 1, pp. 119-139, August 1997. doi:10.1006/jcss.1997.1504.
- [9] Hoashi, H., Joutou, T., and Yanai, K. 2010. "Image recognition of 85 food categories by feature fusion." In *Proceedings of IEEE International Symposium on Multimedia*, pp. 296-301, Taichung, Dec. 2010, IEEE. ISBN 978-1-4244-8672-4.
- [10] Yang, S., Chen, M., Pomerleau, D., and Sukthankar, R. "Food recognition using statistics of pairwise local features." In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2249-2256, 13-18 June, 2010, ISSN 1063-6919.
- [11] Shotton, J., Johnson, M., and Cipolla, R. "Semantic textron forests for image categorization and segmentation." In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1063-6919, 23-28 June, 2008, ISSN 1063-6919.
- [12] Wang, D.H., Kogashiwa, M., and Kira, S. "Development of a new instrument for evaluating individuals' dietary intakes." *Journal of American Diabetes Association*, 106(10), pp. 1588-1593, Oct. 2006.
- [13] Wengreen, H.J., Madden G.J., Aguilar S.S., Smits R.R., Jones B.A. "Incentivizing children's fruit and vegetable consumption : results of a United States pilot study of the Food Dudes program." *Journal of Nutrition Education Behavior*. 2013;45(1):54-9.
- [14] Martin C.K., Han H., Coulon S.M., Allen H.R., Champagne C.M., Anton S.D. "A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method." *British Journal of Nutrition*. 2009;101:446-456.

# Evaluation of Mobile Learning System for Healthcare Support

Toshiyuki MAEDA<sup>1</sup>, Yae FUKUSHIGE<sup>1</sup>, Takeshi MATSUDA<sup>2</sup>, and Masumi YAJIMA<sup>3</sup>

<sup>1</sup> Faculty of Management Information, Hannan University, Japan

<sup>2</sup> Faculty of Comprehensive Informatics, Shizuoka Institute of Science and Technology, Japan

<sup>3</sup> Faculty of Economics, Meikai University, Japan, Japan

**Abstract**—*This paper presents an evaluation of the mobile learning system for healthcare support, which is designed and developed from view of previous research results and problems. The system is an e-mail text-based communication system independent with sorts of terminals, based on mailing-list system and approved various communications by one-stop account. The system offers users' availability for checking their own data which is reflected by every day's reply data, and daily mails have many variation because they are not automatically sent but nurses writes and sends every day, which let e-mails not routinely but humane and improved. We have validated the system by the field test, where we have evaluated and discussed the results.*

**Keywords:** healthcare support, mobile system, e-mail communication, system evaluation

## 1. Introduction

Health is essential for us human being to live a happy and pleasant life. We need to sustain the motivation to keep healthy habits. Research data by Japanese Ministry of Health, Labor and Welfare address that health consciousness of the Japanese people is getting higher and higher in these days. At the same time, however, the number of people who claim their health problems is increasing. This may be because life styles of Japanese people change and that cause of the irregular life cycle and increase of stress. That indicates many people have any health problem caused from lack of normal nutrition or activity, though high health consciousness. We insist that it is important to have adequate health consciousness.

[1] reports that university students who have high health consciousness make good health action. Furthermore, health problems for aged people should be more severe. We believe people should raise health consciousness in younger ages for preventing health problem. Overweight may not only causes into health disorder, but also affects future life style, and so it is very important to care for weight. We focus on healthcare education as preventing health disorder of student youth and improving "health literacy."

There are some researches for healthcare support systems. [2] researches and develops healthcare system in universities using Web and mobile system. [3] studies calories management for healthcare support using mobile phone applications.

Above systems, however, have problems for our targets as below;

- System complexity for various carriers,
- Lack of health consciousness for nutrition, and
- Lack of communication between students and health office.

System complexity for various carriers causes from Japan-specific matters. In Japan, there are some carriers, which have many equipment types and no compatibilities for contents service. For that reason, we have to pay tremendous cost if we support all equipment. We thus introduce a generic mail-based communication system that can be used by almost all mobile phone terminals.

Furthermore, our system supports not only mobile phone terminals but also other equipments (PC, PDA, and so on) which can send/receive mails, and that leads various approaches by various equipments. We can operate the system only by normal mails, and that means users should not learn additionally how to use, and so the system can be regarded as easy-to-use, lightweight, and generic.

We expect so called "Recording diet method" effect by sending plain-text e-mails written by students themselves. The method is to lead dietary effect for users by recording their daily meal data, and then reinforcing consciousness of caloric intake. In Japan, "Recording diet method" is one of popular diet methods. For instance, [4] researches a recording system of caloric intake by taking and saving each meal image. [5] develops a Web-based "Recording diet method" system and examines field tests. This system expects to improve casual human communication by forcing students to describe health management data and to send mails. [6] explains that students who have moderate dietary and physical exercise habits tend to be tolerant to stress and emotional control, and interventional supports for personal stress and lifestyle is a effective method to train self-management for healthcare activity.

We have been researching learning education support systems[7], [8]. We have then applied and improved these researches into the healthcare system for aiming at innovated methodological system, and have had some field tests[9]. This paper addresses our new healthcare system for communication healthcare, which have solved some problems, which are indicated in Aimos field tests. We have had field tests by the new system. We discuss these matters, and then attempt interventional support to students by frequent

communication. We express features and effects of our system, and discuss field test results as follows.

## 2. Problem for Previous System

We had researched and had a field test for above health consciousness by Aimos. On that system, users send e-mails with everyday's caloric intake and amount of physical exercise to a medical office, and the data are stored in a database on a server. Medical office staffs (including nurses) advise each user adequately using the database. The system is e-mail based communication system, and then all of users (not only students but medical staffs) communicate to each other by only emails. We evaluate the system by some field tests and find the following problems;

- Lack of consideration for user interface as prompting health care support
- No function of accessing user's own data
- Gradual decrease of motivation of students

As Aimos was originally developed for lecture support system, WUI (Web User Interface) for medical office is designed mainly for checking of lectures and results of marking point and so on. For instance, result data are presented mainly with each student, though medical staffs need to review, weight data for all students at a glance. As above, WUI should be improved for medical office use. Students cannot access to their own data such as previous weight records, advices from medical office, and so on.

Furthermore, we find obvious fall of reply rate during the examination period in the field test, and that implies the lowering of motivation for students. We regard retention of motivation as essence for replying everyday's questionnaire and maintaining healthcare.

As discussed above, we designed originally for healthcare support from scratch, and develop a new system which is essentially different from Aimos, improved from lecture support system. The new system supports flexible communication on Web and/or e-mail hybrid methods for sending information, though Aimos supports only Web interface for sending/receiving questionnaires. Furthermore, WUI is improved as medical office's requests such as showing questionnaires replies, and so on.

User interface for students is not changed, where the new system receives questionnaire e-mails from students every day. The questionnaire flow is as follows: First, the system sends an e-mail, which consists of questionnaires, to each student everyday. The questionnaire prompts to reply caloric intake, weight, amount of physical exercise, and so on. Then each student reply with quotation of questionnaire texts, and the replies are automatically analyzed and stored by the system server.

## 3. Features of New System

Our new system is designed and developed as follows (see Figure 1);

- E-mail text-based communication independent with sorts of terminals
- Based on mailing-list system and approved various communications by one-stop account
- Users' availability for checking their own data which is reflected by every day's reply data
- Daily mail variation which is not automatically sent but nurses writes and sends every day, which let e-mails not routinely but humane

As mentioned above, each mail to students are not automatically delivered but written and sent manually by nurses, and that let mails human-like, which may avoid boring with the system. At the same time, we try to send personal mails more frequently. For instance, we have short advices to students who do not reply frequently, or have irregular meals. Furthermore, we modify the questionnaire; For instance, the last item "Tell the good thing of the day" is changed to a vague question "Tell any message or just a tweet" which is easier for students to write down.

We have field tests for students. The tests consist of below:

- Firstly, we collect health data (weight, caloric intake, physical exercise, etc.) through e-mail communication
- Secondly, the server analyzes the data and classified into the database.
- Thirdly, the medical office checks the database and makes some advices respectively if required

## 4. Results and Discussion

We have had a field test and the results are very interested. In this test time, we regarded as test subjects who reply more than two times, including students of very few replies during the test period.

Table 1: Field test statistics.

Number of subjects		25
Return rate	Average	31%
	Max.	59%
	Min.	15%
Weight up/down rate	Average	-0.29kg
	Max.	-5kg
	Min.	+3kg
Numbers of input characters (all)	Average	58.5
	Max.	989
	Min.	2
Numbers of input characters (as for free format)	Average	18.9
	Max.	907
	Min.	0
Free format ratio in questionnaire (Average)		87%

In this test, there are few students who reduce weight very much. Reply rate is, however, much raised rather than our previous test (see Figure 2). This is quite interested because this test period includes some holidays, which may decrease motivation of students.

Weight change of students varies from -5kg to +3kg. In general, students who frequently reply seem to reduce



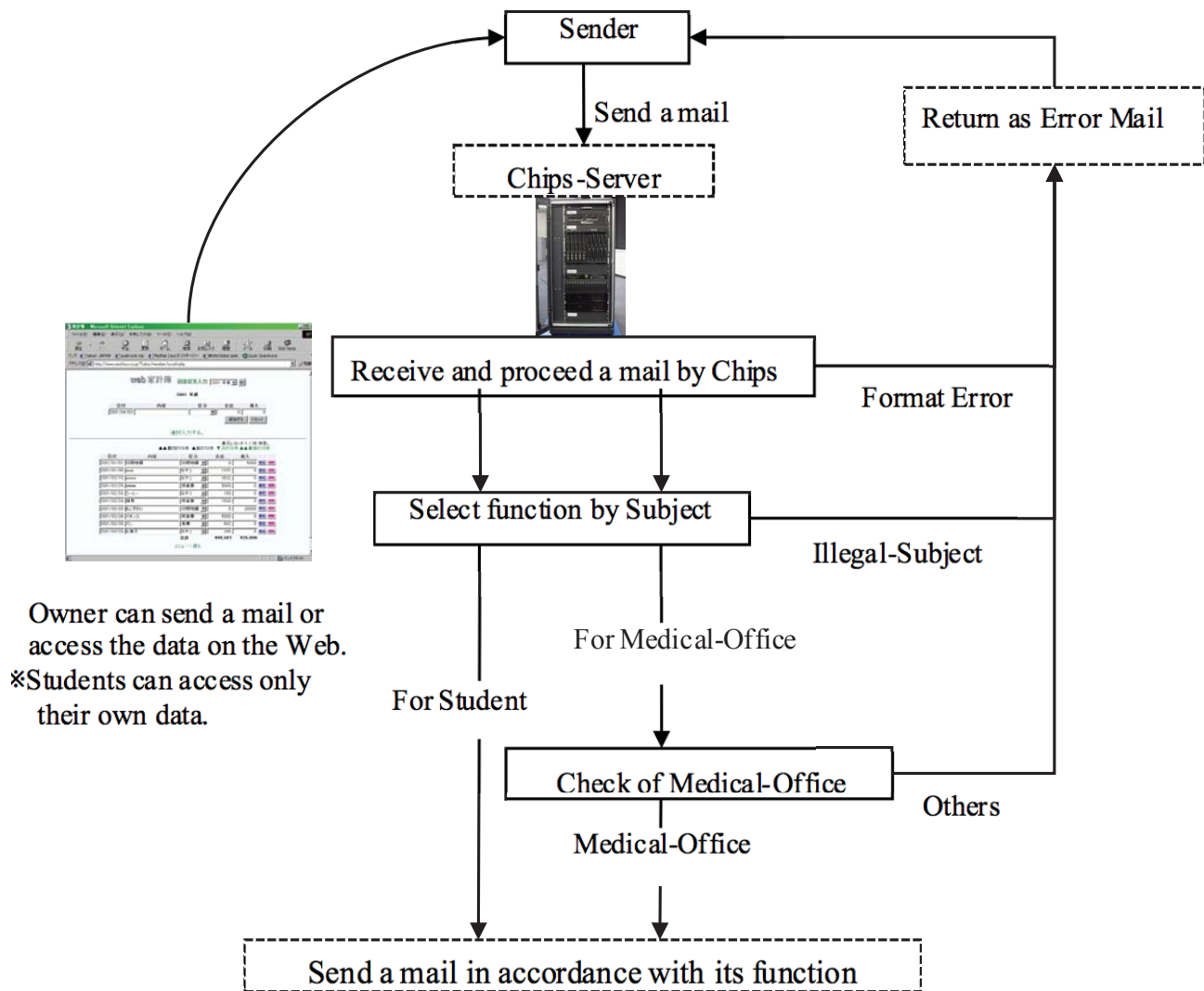


Fig. 1: System flow.

weight. Moreover, according to hearings after the field test, even students who rarely reply do not ignore e-mails from medical office but just read those. Those imply human-written mails promote to read mails, and by those mails, students feel tight connection to the medical office and so they may be relieved. We believe that is a main reason for improving reply rate, as students evaluate e-mails fairly good.

One sentence message included at the last part of the questionnaires filled about 87% of all mails. Among those, quite many mails are not concerned with dieting but just with student life. We assume the message communication is

dealt with a kind of mental healthcare matters, though it is not sure.

Furthermore, we investigate the hearings by text-mining method. We introduce correspondence analysis[10] combined with Japanese token analysis. (see Figure 3)

For analysis purpose, we categorize students into 3 class:

- “minus” : students who reduce weight over 2kg.
- “fair” : students who reduce weight under 2kg.
- “plus” : students who gain weight.

This figure shows some correspondence with “minus” students with the words “Yaseru” (reduce weight), and with “plus” students with the words “Shokuji” (meal), though

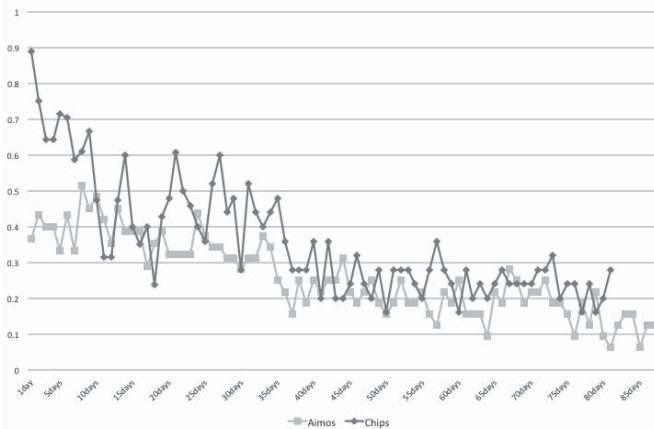


Fig. 2: Return Rate.

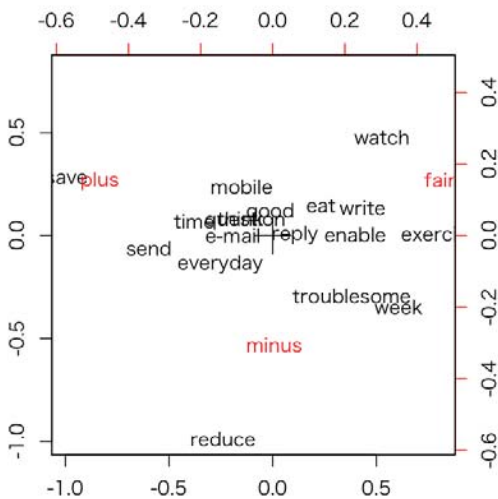


Fig. 3: Correspondence Analysis for Questionnaire.

those classification is not so clear. We have to investigate more precisely.

### 5. Conclusion

We have developed a new system and evaluate the system by the field test, where we have many important data and discussion. In the near future, we analyze the results more precisely, including data-mining methods. We may find the nature properties of rules and then improve the system. For instance, an inspiration mail is sent to who does not reduce weight well, or does not reply so frequently, and the mail may enforce consciousness to the student.

### Acknowledgement

Part of this research was supported by JSPS KAKENHI Grant Number 24520173, 25380630, and 15K03802. This research was also partially conducted under a sponsorship of Grant of Institute of Industrial and Economic Research, Hannan University. Preceding field tests is collaborated with Medical Office at Hannan University. The authors greatly appreciates the supports.

### References

- [1] K. Fujisawa and S. Watanabe, "The Investigation of Health Consciousness and Behavior of Students: A Case of Certain Humanities Students at a Private University," *Bulletin of Institute of Health and Sports Sciences, University of Tsukuba (in Japanese)*, pp. 81–89, 2004.
- [2] W. Y. Jen, "Mobile Healthcare Services in School-Based Health Center," *International Journal of Medical Informatics*, vol. 78, pp. 425–434, 2009.
- [3] C. C. Tsai, G. Lee, F. Raab, G. J. Norman, T. Sohn, W. G. Griswold, and K. Patrick, "Usability and Feasibility of PmEB: A Mobile Phone Application for Monitoring Real Time Caloric Balance," *Mobile Network Application*, vol. 12, pp. 173–184, 2007.
- [4] K. Kitamura, T. Yamasaki, and K. Aizawa, "Food Logging and Processing: Analysis of Food Image," *Journal of The Institute of Image Information and Television Engineers (In Japanese)*, vol. 63, no. 3, pp. 376–379, 2009.
- [5] S. Karikome and A. Fujii, "Analyzing Diet and Access Logs in A Dietary Habit Support System," in *DEIM Forum 2010*, 2010, p. A4.
- [6] K. Takahashi, "A Study of Psychological Factors Related to Health Attitudes and Lifestyles of Japanese Undergraduate Students from View Points of Emotional Responses, Coping Behaviors and Health Locus of Control," *Insurance Management Summary, University of Hirosaki (In Japanese)*, vol. 30, pp. 14–21, 2009.
- [7] T. Maeda, T. Okamoto, Y. Fukushige, and T. Asada, "Learning Session Management With E-mail Communication," in *Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2008)*, Vienna (Austria), 2008, pp. 1787–1792.
- [8] T. Maeda, T. Okamoto, Y. Fukushige, and T. Asada, "Mobile Communication System for Health Education," in *Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2009)*, Honolulu (HI, USA), 2009, pp. 1156–1161.
- [9] T. Maeda, Y. Ando, Y. Fukushige, and T. Asada, "Mobile Communication Effects for Health Care Education," in *Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2010)*, Toronto (Canada), 2010, pp. 556–560.
- [10] H. O. Hirschfeld, "A Connection Between Correlation and Contingency," *Proceedings of Cambridge Philosophical Society*, vol. 31, pp. 520–524, 1935.

# An Integrated Healthcare System and Services for Empowering Patients

Ah-Lian Kor

School of Computing, Creative Technologies, and Engineering, Leeds Beckett University, Leeds, UK  
A.Kor@leedsbeckett.ac.uk

Colin Pattinson

School of Computing, Creative Technologies, and Engineering, Leeds Beckett University, Leeds, UK  
C.Pattinson@leedsbeckett.ac.uk

Richard Pope

Dynamic Health Systems, Bradford, UK  
Richard.Pope@dynamichealthsystems.co.uk

Axel Schulte

Dynamic Health Systems, Bradford, UK  
Axel.Schulte@dynamichealthsystems.co.uk

**Abstract**— The main goal of this proposed research is to radically transform health services with incrementally evolving organisational change. This is made possible by exploiting advanced cloud-based ICT systems to empower patients to self-manage their own health and wellbeing. The project innovation is not the base technology (which is highly innovative by itself) but the rapid continuous improvement of technology enabled service delivery (i.e. following the PDSA iterative cycles) based on continuous user and technical requirements analyses. A range of ICT-enabled services are provided to all citizens with health records (including long term and short term patients) at greatly reduced costs. They are: a customisable e-health passport which contains all essential health and wellbeing information about the patient; intelligent data analytics and decision support system; clinical team view (with health intelligence facility); hybrid diagnostic expert system for teleconsultation; and online Community of Support.

**Keywords**—healthcare, system, services, agile, self management, self empowerment, patient, ICT, cloud, pdsa, biomimicry, value chain, patient journey

## I. INTRODUCTION

According to the European Commission, Europe is already spending nearly 10% GDP on healthcare due to EU ageing population[1]. An increasing communicable and non-communicable disease burden and the economic challenges jeopardise the sustainability and equity of European health and care systems (ibid). The EU Health2020[2] aims to: “significantly improve the health and well-being of populations, reduce health inequalities, strengthen public health and ensure people-centred health systems that are universal, equitable, sustainable and of high quality”.

In order to support the EU Health2020 priorities and be aligned to the EU Action Plan on eHealth[3], this paper proposes an inclusive, integrated and user-centric ICT system and services to assist prevention, diagnosis, prognosis (and prediction), health monitoring, treatment, and lifestyle management. The name of the proposed system is *Integrated Teleconsultation, Telehealth, and Telecare System and Services (IT3SS)*. Its innovation is a **radical transformation of healthcare system and services** accompanied with an **incremental evolution of the organisation**. It will be scalable due to its modular and layered cloud-based architecture (see Figure 5) that is made possible with plugins (**sustainable**), **patient centred** system and services. IT3SS provides the following functionalities: (i) empowerment for patients to participate in decision making in the health service and co-creators[4] of health (which prioritises patient centredness and patient involvement, as well as self-management interventions); (ii) continuity of non-emergency self-care; (iv) accessibility to own health information for better care of their health; (v) health intelligence and decision-support for clinicians; (vi) use of social media to provide support. The radical innovation is a shift from a system-centric to a patient-

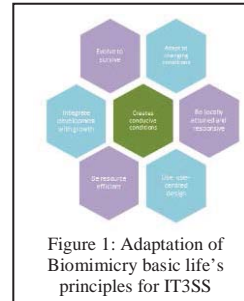
centric paradigm where the dataflow is about the patient and back to the patient in real time if necessary. It will include the following: data integration (including syntax and semantics) between data from health telemonitoring devices and the patient health record system data; an intelligent data analytics and decision support system module that will provide appropriate processed information to the patients and health intelligence as well as recommendations on co-morbidities/poly-pharmacy to clinical teams; hybrid expert system for teleconsultation; and Online Community of Support.

## II. UNDERLYING CONCEPTS AND APPROACHES

The IT3SS system will be anchored on an adaptation of biomimicry basic life’s principles[5] (see Figure 1) and Porter’s value chain model (see Figure 2).

### A. Adapted Biomimicry Life’s Principles

Biomimicry is an emerging approach to innovation that seeks sustainable solutions to human challenges by emulating nature’s time-tested patterns and strategies and its goal is to create sustainable products, or processes[6]. The ultimate goal for the IT3SS is to **create conducive conditions** (based on biomimicry life’s principles depicted in Figure 1) for the development and deployment of an ICT-enabled healthcare system and



services which is sustainable, efficient, effective and beneficial to patients and the clinical team. The principles are as follows: (i) **Evolve to survive**: replicate strategies that work (note: IT3SS will be built on Vitrucare and Wellness Layers’s success stories on e-health care); (ii) **Be resource efficient**: use multi-functional design (as depicted in Figure 5, IT3SS will afford multi-functionalities to support patients in their self-empowered journey); (iii) **Use user-centred design** (primary users (patients and clinical team) will be involved throughout the lifecycle of IT3SS health care system and services); (iv) **Be locally attuned and responsive**: use feedback loops and leverage cyclic processes (see the agile methodology and PDSA cycle discussed in Section 5), cultivate cooperative relationships (amongst the patients, clinical and non-clinical teams in IT3SS), use readily available materials (e.g. use and extend Vitrucare and Wellness Layer’s existing platform); (v) **Adapt to changing conditions**: incorporate diversity (e.g. diversity of supporting activities shown in Figure 2), embody resilience through variation redundancy and decentralisation (note: GPs will only be responsible for their local patients), maintain integrity through self-renewal (note: this is made possible via continuous improvement and updates); (vi) **Integrate development with growth**: build form from the

bottom-up (note: the IT3SS healthcare system and services will be built based on collated user requirements and analysis), combine modular and nested components (see Figure 5 for IT3SS modular cross-layered architecture).

*B. Adapted Porter's Value Chain Model (1985)*

The term "Value Chain" was introduced by Michael Porter (1985) in his book "Competitive Advantage: Creating and Sustaining Superior Performance"[7]. It is based on the process view of an organisation seeing an organisation as a system with sub-systems (each with inputs, transformation processes and outputs)[8]. The manner the value chain activities are carried out determines the costs and affects the profits. On the other hand, in IT3SS, the value chain provides a process view of an integrated health care system and services. The value chain activities are patient-centred and the beneficiaries will be patients, clinical team, etc... The aim of

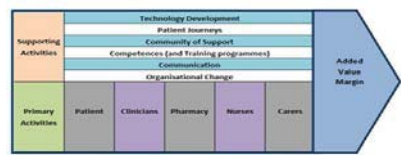


Figure 2: IT3SS Value Chain Model for ICT-enabled Healthcare and Services

IT3SS value chain model is to improve patients health and wellbeing via empowerment, provide integrated patient centred information (as a result of analysed data) which is accessible by a team of healthcare professionals, reduce costs (i.e. reduce added value margin in Figure 3), effect organisational change (in terms of redefinition of roles and responsibilities within the management/operational structure).

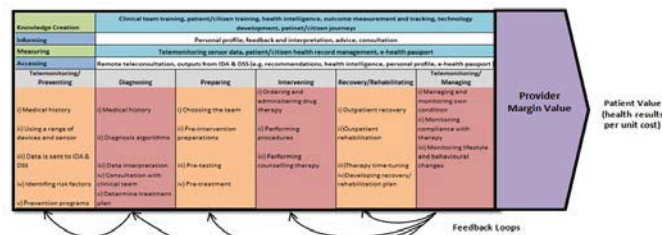
IT3SS value chain model in Figure 2 encompasses components: (i) **actors** (i.e. patient, clinical team (clinicians, pharmacy, nurses, etc...), and carers if appropriate); (ii) **primary activities** of the actors (e.g. self telemonitoring, engagement, consultation, giving advice, act on health intelligence, etc..., see Figure 3 for a list of appropriate activities); (iii) **means of delivery** for the ICT-enabled healthcare and services (technology development, patient journeys, Community of Support, Competencies Building (training programmes), communication between the patient and clinical team, and organisational change); (iv) **supporting activities** for (iii). According to Porter, it is imperative that primary activities be closely linked to support activities in order to help to improve their effectiveness or efficiency. The added value margin of IT3SS depends on the seamless linkages between all activities in the value chain. Patient Journey in the Healthcare System and Services (An adaptation of Porter's Care Delivery Value (CDVC) Chain Model)[9] According to Porter and colleagues (2009)[10], the care delivery value chain is a framework that allows systemic analysis of value creation across the myriad of activities that occur during a patient's health care lifecycle. Figure 3 shows an adaptation of Porter and Teisberg (2006)'s [11] model for IT3SS ICT-enabled health care system and services.

Value-based healthcare and services delivery requires a radical transformation (note: this is IT3SS's innovation) because in this new paradigm, the patients are active and proactive drivers of their own health care and wellbeing, supported by other relevant actors (as shown in Figure 2). The three important care transformation enablers[12] relevant to IT3SS context are: (i) the health care system and services **value chain** (shown in Figures 2 and 3) which involves a systematic approach to process identification and analysis which we call patient journeys (note: Figure 4 depicts the teleconsultation non-emergency care journey while the following list of care-related processes (shown in Figure 3): telemonitoring /preventing, diagnosing, preparing,

intervening, recovery /rehabilitation, telemonitoring /managing is a generic example of a self-empowered /managed health journey in IT3SS; (ii) **ICT** provides a means to transform telemonitoring sensor data and patient records into useful information that will displayed in e-health passports (that will be shared seamlessly between the patients and clinical teams), health intelligence (for the clinical team to inform policy, action plans, etc...). The cloud-based ICT infrastructure employed for the deployment of IT3SS system and services is depicted in Figure 5; (iii) systematic processes for **knowledge creation** to support continuous improvements in health care and services and training. Results (patient health outcomes and costs) are iteratively measured, analysed, and reported (note: the iterations are shown in the feedback loops in Figure 7).

III. RELATED WORK

The European Code of Practice for Telehealth Services is found here[13]. The European Code of Practice for telehealth services is closely aligned to the European Commission's eHealth Action Plan 2012-2020[14] and its goal is to provide a robust high level quality benchmark. The code of practice underpins telehealth (and telecare) services which promotes preventive health, public health, healthy lifestyle and wellbeing. It addresses the needs of service users, service providers, carers, and provides a framework for service providers to plan and manage their services in appropriate inclusive and ethical ways. In addition, the code points to skills, knowledge and competency requirements for service staff. Telehealth services are provided through diverse range of technologies[15]: interactive simple telephony, interactive TVs, webcams, video links, wired or wireless telecommunication and computing devices (e.g. smartphones, tablets, etc...). Other relevant technologies are: sensors; environmental controllers; apps which facilitates secured sharing of information relating to health, wellbeing and relevant activities; social media. VitruCare is a web-based service (accessible via a range of devices – PC, laptop, tablet, smartphone) provided by Dynamic Health System[16] and prescribed to patients by their General Practitioners (GPs). The initial deployment of VitruCare caters for patients with



Note: IDA & DSS – Intelligent Data Analytics and Decision Support system  
Figure 3: An Adaptation of Porter and Teisberg (2006)'s model for IT3SS health care system and services

long term conditions: Diabetes Type I and Type II, Hypertension, Chronic Heart Failure, Asthma and Chronic Obstructive Pulmonary Disease. However, as previously discussed, the VitruCare cloud-based platform is scalable and thus could be extended to add more diseases. VitruCare is patient focused, and therefore is designed to deal with multi-morbidity, providing coherent guidance to people living with more than one condition. One of the services provided by VitruCare is help patients (with GP support) make practical plans to manage their health, track their progress in real-time, change their lifestyle and improve their overall health. The VitruCare approach sees the patient going on a new journey of supported self-care, with signposts provided by the care team. Positive feedback[17] has been given by patients regarding the

use of VitruCare. GPs can view patients' clinical data taken directly from the secured electronic health record and communication between clinicians and patients is also secured. The VitruCare approach complies with the stringent information governance requirements of the NHS (outlined by Information Commissioner's Office[18]).

Here, we shall review how Orange[19] supports the Digital Hospital Programme[20] in France which aims to provide access to health care by everyone, guarantee continuity in providing care services between hospitals and non-hospital practices, and improve medical practices and governance in both public and private healthcare facilities. Orange provides digital tools for mobile services that revolutionises access to healthcare and medical practices through sharing of medical data. In this innovation, the system is patient-centric who benefits from changes in treatment, follow-up and quality of life. Thus, it is imperative that the Hospital 2.0 concept focuses on accessibility for users, interactivity of the care process and simplification of routines for healthcare personnel[21]. Orange has rolled out a range of innovations: access to services via multimedia terminals, interactive TVs, webcams, intranet access; emergency medical response services equipped with on-board IT systems that enable geolocation and automatic transmission of data to patient files; setting up of a health hub for training; medical monitoring within the home or office setting; m-health[22]: use of mobile technologies (phones, laptops, tablets, PDAs, etc.) to provide healthcare solutions.

In 2008, Sweden[23] has more than 100 applications for telemedicine services, for example, teleconsultation and videoconferences (doctor-to-patient); telemonitoring. The service for remote patient health monitoring was launched in 2005 using GPRS mobile technology which enables hospital and healthcare staff to remotely monitor patients with sickness or diseases that do not require hospital treatment or other types of medical care. Patients could also access personalised services (e.g. bookings, access own record, etc...) and relevant health information on the web via e-Health Portals (e.g. waiting times, practical advice on self-care, etc...). In Sweden the "Healthcare-digital-network" service offers secured transmission of patient information. Other innovations under consideration are: all drug prescribers would have access to the patient's entire drug history and appropriate prescriber support; the National Patient Summary shall be fully deployed throughout the healthcare services and in the relevant parts of the social services.

Currently, more than three thirds of health care centres in Finland[24] seem to use eReferrals (an e-message written by a GP for referring a patient to the hospital) and accept eDischarge letters. Digital patient data is stored in each individual health care institutions and thus secured data transfer is necessary. Currently, more than half of the 21 district hospitals provide e-consultation services. In 2007, about 80% of the hospital districts deliver teleradiology services while approximately 90% of the hospital districts have some form of e-distribution of laboratory results (known as telaboratory). The national ICT infrastructure provides the following functionalities: a national eArchive in which all EHRs are available online with patient consent; a national ePrescription database; an eView for citizens to access to their own patient data and log data.

There are 429 municipalities in Norway[25] and 75% of them use the electronic health record system to support nurses and GPs in care communities. E-services delivered are: e-prescriptions (can be shared among the clinical team members,

and co-ordinated at the national level); self-care services accessed through the National Health Portal (Helsenorge.no). The functionalities offered through this portal are: (i) Self Empowerment - self health management: repeat prescriptions, appointment booking; eConsultation with GPs and other health care professionals; mHealth – self assessment, trackers, etc...;(ii) Health data access – access to health record and log; register for vaccination; (iii) access to health and illness information; (iv) access to other healthcare services – provide feedback; locate healthcare and care providers, contact the web portal host, user rights.

Two major health service challenges in Ukraine[26] are: low quality of healthcare and unequal access to health services. In order to close the gap, Ukrainian specialists adopt the telemedicine and teledermatology technologies. In 2012, 820 patients underwent remote melanoma screening using telediagnostic technologies and a remote dermatology consultation center was set up in the Donetsk region as a pilot project. According to experts, the sensitivity of telediagnosis was 98.8%. In 2010, a telemedicine network with 6 nodal stations was set up and further plans were made to extend and expand the implementation of telemedicine across the nation.

Wellness Layers[27] has a head office in the USA but with an International Research and Development Office in Israel[28]. They leverage their platforms (details of the platforms are found here[29]) to provide enterprise-level solutions which are flexible, customisable for a mix of target population: citizens, patients, and health care professionals. These platforms facilitate interactive environments to engage the users, guide, educate and connect them in order to address their health and wellness needs in a personalised and supported manner. Each of their portal solutions includes interconnected applications from the three layers of engagement that are called me-we-info where the Me Layer aims to guide each user to his/her own path to achieving goals and optimal health and wellness via smart planners (note: these are accessible by coaches) and trackers. As for the We Layer, it connects members to achieve better outcomes with support and guidance of other members with similar goals and conditions. Wellness Layers interconnected social applications are linked to the Me Layer to support each individual's plan

(with privacy), specific groups' needs, segmented discussion forums and personal pages to share stories, struggles and achievements. The Info Layer newsfeed based informative and

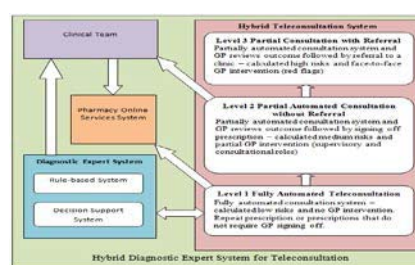


Figure 1. Hybrid Diagnostic Expert System for Teleconsultation

educational layer that transforms a static informational site into a personalized and ever-changing destination with dynamic feeds and calls to action. It also includes search and directories with smart filters. Further details for the me-we-info layers can be found here[30].

#### IV. ARCHITECTURE FOR IT3SS SYSTEM AND SERVICES

IT3SS integrated software architecture (depicted in Figure 5) is modular and cross-layered (just like the Wellness Layers's Platform)[31]. The design and development of each module in the IT3SS system and services will be based on the component-based software design and development approach[32] with the following principles: (i) components are independent (note all the IT3SS modules can be developed

independently at the initial stage; (ii) communication is through well-defined interfaces (note: in IT3SS, the interaction amongst the modules are made possible via integrations; (iii) components have shared platforms (note: in IT3SS, a cloud-based platform will be used). The architecture consists of 4 layers (i.e. Data Layer, Platform Layer (Data Access layer and Integration Layer), Application Layer, and Access Layer) and 5 modules (Secured Individual Patient Personalised View with e-health passport, Intelligent Data Analytics and Decision Support System, Secured Clinical Team View, Hybrid Diagnostic Expert System for Teleconsultation, and Community of Support). The following are the components in IT3SS health care system and services.

**A. Data Layer**

The data layer will comprise data from two sources: telemonitoring data in the home or community setting (i.e. vital signs for health management, activities and fitness tracking for wellbeing) and patient health records. Telemonitoring of vital signs from the home is an important aspect of telehealth services because it is necessary for the management of patients with chronic conditions (e.g.

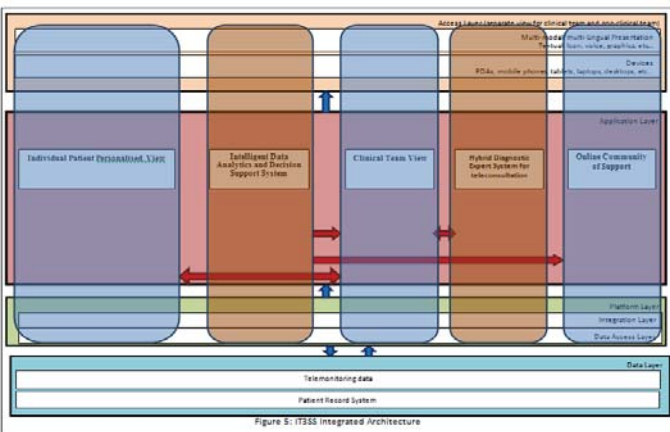


Figure 5: IT3SS Integrated Architecture

congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), or diabetes). Currently, commercial telehealth monitoring systems in the market are very expensive and rendered unaffordable by most patients. Integrated telemonitoring systems such as the Honeywell HomeMed Telemonitoring System[33] (e.g. Sentry telehealth monitor, or the Genesis telemonitor) will be beyond the reach of many patients. A patient using Honeywell HomMed's Sentry Telehealth Monitor[34] can collect and securely transmit data (via Honeywell HomeMed's patented wireless pager system) on five key vital signs (heart rate, blood pressure, oxygen saturation, temperature, and weight) to the Central Station Diagnostic Software. Patients are also guided through a questionnaire (responses: yes or no) for a thorough health assessment.

IT3SS aims to reduce unequal access to health services. This is achievable with the provision of open-sourced systems and services accompanied by affordable telemonitoring devices that could be employed within the home or community setting. An intelligent data analytics and decision support system will provide automated monitoring of the patients' health conditions and alert the patients themselves and clinical team appropriately. It will also provide a list of information for e-health passport in the individual patient personalised view (see Figure 5) that is also accessible by the appropriate clinical team. IT3SS web portal will provide access to its system and services via a range of devices. Activities and fitness tracking will be made possible with

wearable technologies. This will be part of the patients' personalised services and interventions for healthy lifestyle as well as personalised programmes for health management are found in the e-health passport.

**B. Platform Layer**

The existing Vitrucare platform is cloud-based (e.g. Microsoft Azure). Additionally, cloud services offered via the Vitrucare platform are scalable and could support a large number of concurrent users. The Platform Layer has two sub-layers: Integration Layer and Data Access Layer. Cloud Integration Platform as a Service (iPaaS)[35] is an emerging facility for integrating cloud applications with one another and it should be sufficiently open for integration solutions to be built and customised. However, developers have to be familiar with tools and processes to build and configure integration-style applications.

**Data Access Layer**

This layer interfaces with the patient record system and telemonitoring data. This layer will ensure the compliance to the Data Protection Act (DPA)[36] and EU General Data Protection Regulation (GDPR)[37]. Microsoft Azure has developed an extensible compliance framework[38] for the design and building of services that are compliant to a diverse set of regulations. Some of the specific compliance programmes are: ISO 27001[39] (information security management); UK Cloud Security Principles[40]; etc...

**Interoperability of Systems**

The Health Level 7 (HL7)[41] provides a framework (and related standards) for the integration, sharing, and management of electronic health information. These standards define the following (which is necessary for seamless integration between systems): mode of communication, setting the language, structure and data types. In addition to this, HL7 manages interoperability standards for healthcare information[42].

**C. Application Layer**

**Teleconsultation (via hybrid diagnostic expert system)**

Teleconsultation[43] typically involves consultations with specialists at a distance, with the patient being examined or otherwise somewhere in the system. In recent development (e.g. in the US), teleconsultations are held via secure networks, Skype, or other video conferencing facility. Service models which incorporate Teleconsultation can reduce health inequalities by bridging geographical remoteness[44] because

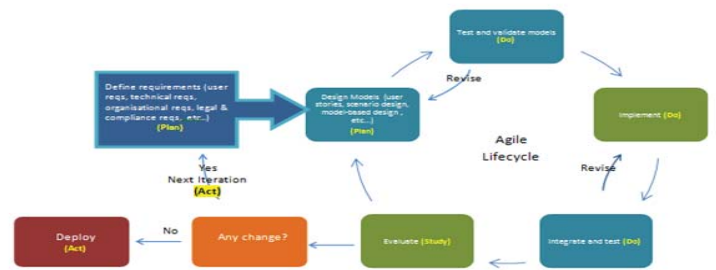


Figure 7: Integrated Agile Lifecycle and PDSA

patients could access the services within the home setting or in a community hub. In the UK, the Airedale NHS Foundation Trust has developed a set top box based video communication system that works through an interactive TV and domestic broadband (ibid). Outpatient teleconsultations could occur between a patient (in a home setting) and clinician (in a clinical environment). Examples of such services are: UK - NHS 111 (fast-tracked non-urgent medical service)[45];

Sweden – 1177 (website and helpline for health services)[46]. In the IT3SS project, a hybrid diagnostic expert system for teleconsultation (shown in Figure 4) will be developed so as to reduce unequal access to health services and also to supplement NHS 111 and Sweden 1177. It will link the patient to the clinical team and online pharmacy via the expert system. As previously discussed, **our novel hybrid teleconsultation system** has 3 levels: (i) Level 1: fully automated teleconsultation where prescriptions do not require GP signing off; (ii) Level 2: partial automated consultation without referral where the GP remotely reviews the outcome of the teleconsultation and sign off prescription if necessary; (iii) Level 3: Partial consultation with referral where the GP remotely reviews the outcome of the teleconsultation and refer to a clinic if necessary.

### Intelligent Data Analytics System and Decision Support System (IDA & DSS)

IDA & DSS will have two modules: data analytics and decision support modules. The approach employed for the data analytics module will a data mining approach and the two main data mining models used will be: descriptive models (or profiling models) and predictive models[47]. The former provides insights into the distribution of the data by describing, summarising, organising the data (e.g. patient's personalised profiles, activities and performance tracking in the e-health passport, and health intelligence for the clinical team in IT3SS) and also help uncover interesting patterns within the data (for diagnosis). On the other hand, the latter attempts to provide a prediction or forecast based on drawn conclusions (predicted health/illness for appropriate interventions in the e-health passport). Predictive analytics will help preventive medicine and boost public health because at-risk patients could be identified and early interventions (such as change in lifestyles) could help prevent diseases and reduce risks[48].

### Community of Support

Communities of Practice[49] will form the basis for IT3SS's Community of Support. It will provide an environment for knowledge sharing, communication, sharing trusts, beliefs, learned lessons, insights, anecdotes, and values. Additionally, it gives its members a sense of identity, particularly if the communities are special interest groups or stratified groups based on a list of possible criteria such as illness, health conditions, gender, language preferences, demographics, etc... These criteria could be used for the special interest groups in IT3SS project. IT'SS Community of Support will be an extension of Wellness Layers's existing social community infrastructure and the use of such social media will abide by the practical and ethical guidelines provided by the British Medical Association (BMA)[50].

#### D. Access Layer

The development of the access layer will appropriately address web usability and accessibility. Usability[51] concerns product designs which are effective, efficient, and meet user satisfaction. It is a part of human-computer interaction (HCI) which provides guidelines for designs, for example: (i) 10 Usability heuristics for user interface design by Nielsen[52] (i.e. visibility, mapping of virtual and real world, user control, consistency, error prevention, recognition, flexibility, aesthetics & minimalist design, error recovery, help); (ii) Shneiderman's 8 golden rules of interface design[53] (i.e. consistency, shortcuts, informative feedback, dialogue, simple error handling, easy reversal of actions, internal locus of control, and reduce cognitive load). In order to create positive

user experiences, it is essential to follow a user-centred design approach. Web usability and accessibility are closely related in terms of goals, approaches, and guidelines. Accessibility is to ensure an equivalent user experience for users with disabilities, and age-related impairments and this means that users with disability will be able to perceive, understand, navigate, interact with the web services, and participate equally without barriers[54]. This is made possible with assistive technologies (e.g. audio, screen magnifiers, etc...). The W3C Web Accessibility Initiative (WAI) has developed a set of guidelines that are known as web accessibility standards which include considerations for people with auditory, cognitive, neurological, physical, speech, and visual disabilities, including people with age-related impairments (ibid). It is noted that the majority of health information on the internet is in English and is thus inaccessible to non-English speaking users[55]. However, based on the Equality Act 2010[56], no user of any government service should be excluded on the basis of disability. According to the EU telehealth services code of practice[57], services should be accessible via all kinds of communications technologies: simple telephony, television and web-cams, video links through fixed or wireless telecommunication networks and using devices such as smart phones and tablets.

In summary, the IT3SS Access Layer will be designed and implemented based on the discussed usability and accessibility principles so as to reduce unequal access to health care services. It will follow the user-centred design approach where user requirements are being collected, analysed and fed into the design as well as implementation of IT3SS system and services.

## V. METHODOLOGY

The IT3SS system lifecycle will the following integrated approach/methodology: (i) user-centred design approach; (ii) agile methodology for software development (integrated with PDSA (plan, do, study, act) with iterative cycles); (iii) integrated software architecture.

### A. User-Centred Design Approach[58]

According to Norman, et. al. (1986), there are two sides to a system interface namely: system perspective and user perspective. The system perspective is changed through proper design while the user perspective is changed through training and experience. User experience will be positive if the user has a good conceptual understanding of the system which has been designed and built around the user needs and requirements. Thus, the user-centred design which focuses on the user's needs will entail the following[59]: user requirements analysis, activity/task analysis (necessary for the user interaction); initial testing and evaluation, iterative designs.

### B. Integrated Agile Methodology for Software Development (with PDSA)

The PDSA Cycle (also known as Deming Wheel/Cycle in Figure 6)[60] is a systematic series of steps for continuous improvement of the IT3SS health care system and services. The steps are: (i) **Plan**: identify a goal or purpose, define success metrics, and action plan; (ii) **Do**: implement action plan (e.g. make a product); (iii) **Study**: monitor outcomes, validate the plan, identify areas for improvement; (iv) **Act**: revise step (i) based on results in step (iii). These four steps are iterated as part of a continuous

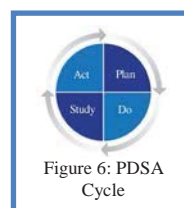


Figure 6: PDSA Cycle

improvement cycle. Agile methodology[61] for software development will be adopted for this project due to its simplicity, flexibility, and adaptability. Just like PDSA, the agile methodology is iterative and incremental, with every aspect of the development lifecycle (i.e. requirements, design, implementation, evaluation, etc...) being continually revisited for continuous improvement. In the agile paradigm, it will provide greater user satisfaction because it delivers a potentially consumable solution for each iteration and enables users to enhance (i.e. evolve for the better) their requirements throughout the entire project[62]. Additionally, it has a quicker delivery compared with the traditional waterfall methodology. Figure 7 depicts our integrated agile methodology and PDSA. The steps involved are: (i) **PLAN:** Define requirements - user requirements: patients, clinical team; technical requirements; organisational requirements; legal and compliance requirements; (ii) **PLAN:** Design models – patient journeys, user stories (i.e. begins with understanding the user, capture it in the form of a persona (a tool to help visualise the user and needs, followed by using the tasks and goals of the “personas” to support decisions on system functionalities)[63]), scenario-based designs (i.e. grounding design in an understanding and capturing the interrelationships of tasks users carry out over time)[64], model-based designs (e.g. Object, View, and Interaction Design, or OVID[65] which systematically builds models using based on a specification of the user model; (iii) **DO:** Test and verify design models in (ii) using simulations or experimental environment; (iv) **DO:** Implement – Develop system and services for the real environment; Intelligent Data Analytics and Decision Support System, Hybrid diagnostic expert system, and Community of Support; (v) **DO:** Integrate and Test – integration within and between the 4 layers: Data Layer, Platform Layer (i.e. Data Access Layer and Cloud Integration Layer), Application Layer, and Access Layer); (vi) **Study:** Evaluation of IT3SS System and Services – expert appraisal, users , impact analyses; (vii) **Act:** Next iteration if revision is required. On the other hand, if the evaluation results meet the IT3SS system and services goals, then, it shall be deployed in the real environment.

VI. INNOVATION POTENTIAL OF IT3SS SYSTEM AND SERVICES

Innovation Potential will consider IT3SS's performance and innovation. Firstly we the types of innovation[66] that emerge from IT3SS are: product and services innovation (i.e. creates new ICT-enabled patient journeys); process innovation (i.e. creates new processes to support the new ICT-enabled services delivery); position innovation which involves the changes in the context in which the product and services are introduced (i.e. changes to management structure, roles, responsibilities and powers); paradigm innovation which involves changes in the underlying mental models which frame what the organisation does (i.e. a change from system-centric paradigm to patient-centric paradigm where patients are self-empowered and self-manage their own conditions supported by technology and the clinical team). The integrated teleconsultation, telehealth, and telecare system and services will effect a radical transformation of patients' roles as co-creations of effective health care services accompanied by merely an

incremental organisational change (i.e. change in management structure (relevant for co-ordination and communication amongst the clinical team members), roles, responsibilities and powers). In order to ensure a successful impact for IT3SS system and services, innovation management is necessary because it identifies 4 vital aspects for enhancement of IT3SS Innovation Potential. The EU e-Health Action Plan 2012-2020 (Innovative healthcare for the 21st century, and a part of the Digital Agenda in Europe 2020 Strategy) provides the potential for IT3SS system and services to be marketed across Europe which makes the ‘reach’ aspect of the impact significant. IT3SS's modular architecture and open-source interoperable technology could also be easily plugged into any existing healthcare system with prior integration work and thus, could be marketed separately. In addition to that, it makes the IT3SS flexible, extendable and scalable. Even though the entire IT3SS target sample could be trained

Beneficiary	Expected Benefit
Patient	Reduction in hospital admissions
	Reduction in A&E visits
	Reduction in home visits
	Reduction in surgery visits
	Reduction in mortality
	Increased access to health care services
	Reduction in waiting times
	Increased flexibility and choice in general practice access
	Increase in quality of life
	Increase in patient satisfaction
	Increase in patient acceptance
	Increase in patient engagement
	Increase in health and wellbeing awareness
	Enhancement of wellbeing health plan
	Behavioural change
	Increased competency in the use of IT3SS system and services
	Increased levels of engagement (i.e. utilisation level) and traffic through IT3SS web portal
	Reduction in demand for acute and out of hours services, A&E services
	Reduction in diagnosis intervention errors
	Extended access to primary care
	Number of timely diagnosis and intervention
	Number of ICT-enabled contacts compared to face-to-face contacts
	Reduction in resources demand
Reduction in demand for face-to-face consultations	
Reduction in costs	
Increased productivity for GPs	
Increased competency in the use of IT3SS system and services	
Reduction in overall demand for a wider healthcare services	
Reduction in the number of severe episodes and complications	
Increased confidence in Intelligent Data Analytics and Decision Support System (IDA & DSS)	
Increased health provider acceptance	
Increased level of safety	
Care efficiency gains	
Cost-effectiveness of system and services (i.e. interventions, etc...)	
Cost minimal gain	
Productivity gain	
Reduced environmental impact of the wider health system	
Increased added value	
Enhanced financial performance of local health economy	
Improved cooperation and secure information exchange among the actors	
Support	Improved interaction amongst patients and health professionals

\*note: the target is provisional and will be revised after each iterated PDSA cycle  
 Table 1: IT3SS Expected Benefits for the Actors

systematically during the system lifecycle, it would be impossible to training the entire patient population and beyond (including other EU countries). Consequently, in order to address this weakness and also to cater for repeated views, online step by step training (e.g. via youtube, etc..). Another weakness to be addressed is the inability to achieve IT3SS system and services full functionality potential if an online patient health record system is not in place. If this happens then the possible solution is that patient records are created in the IT3SS Data Layer via an additional IT3SS secured facility. Resources will be required for the translation of IT33 contents to languages across Europe and the world. Existing products/services in the market that could potentially pose as a threat are: (i) Honeywell Lifecare Solutions[67] (including the HomMed Monitoring System); (ii) Microsoft Health Vault[68] which enables users to gather, store, use and share health information[69]; (iii). iHealth[70] that enables users to store information in the iHealth Cloud and also provide mobile apps to automatically track and manage their key health vitals in one place. As a matter of fact, existing products/services will not pose as serious threat to IT3SS because of the big untapped market share for all providers.

VII. EXPECTED BENEFITS

A. IT3SS Expected Benefits



A list of IT3SS expected benefits for the actors is shown in Table 1 (note: some of these are collated from these sources[71-73]) while a list of general expected benefits is

Dimensions	Expected Benefits
Health	Improved health and wellbeing More effective health promotion and disease prevention. More active lifestyles supported by ICT. Improved disease management. Increased quality of healthcare. Increased safety of healthcare. Development of new healthcare journeys with clear definitions of responsibilities of relevant actors. Improved health information (health informatics) and evidence-based decision making.
Social	Tackle societal challenges Inclusion so as to reduce unequal access to health care and services.
Economics	Industrial leadership and boosting competitiveness. Improved local health economics.
Environmental	Reduce environmental impact (carbon footprint) via reduced travel.
Technical	Innovation in SMEs made possible via knowledge creation. Data integration of a range of sensor data and legacy systems. Standards for data transfer from sensors. Hardware and software integration. Modular and cross-layered architecture. Sustainability, scalability of the technology is due to: cloud-based platform, modular architecture. Increased technical acceptance due to user-centred design and training.
Competence	Training for competency building (clinical and non-clinical teams) in the use of the technology; interpretation of data analysis results and exploitation of evidence-based supported decision making.
Organisational	Effective co-ordination and communication is necessary due to complexity of the system and services (i.e. journeys and involvement of a range of actors). Organisational management structure, definition of roles and responsibilities, as well as timely allocation of well-defined powers (for the journeys and also the co-development of the health/wellbeing plans). Increase organisational performance.
Policy	For ICT-enabled health care and services.
Business Value	A business model for a 'profitable' health care delivery system and services (look at Figures 4 and 6). Increase business efficiency. Increase transactional benefits. Increase number of transactions executed. Increase task productivity. Increase task innovation. Increase net value.
Information	Increase information quality. Increase information access. Increase information flexibility. Increase information security and privacy.

Barrier	Reasons
User barrier	lack of awareness lack of confidence and trust lack of access
Security and Privacy barrier	lack of competence lack of security in transfer of sensor data lack of system security lack of data privacy
Technical barrier	lack of interoperability lack of legal clarity lack of transparency regarding the utilisation of data collected inequality in access to services due to geographical differences inequality in access due to social deprivation lack of availability of user-friendly tools and services poor adoption of standards in eHealth systems.
Standard Barrier	data exchange Communication standards for data interoperability Interoperability design guidelines for personal health systems Multimedia e-health data exchange services Service and capability requirements for e-health monitoring services
Legal and Regulatory barrier	data protection (protect rights) inadequate or fragmented legal frameworks
Capital Investment and Benefits Barrier	lack of legal clarity around m-health lack of evidence-based cost effectiveness benefit limited large-scale evidence of the cost-effectiveness high start-up costs involved in setting up eHealth systems
Organisational barrier	lack of Organisational interoperability
Market barrier	problem with market fragmentation
Diffusion and adoption barrier	low level of diffusion and adoption

depicted in Table 3. The expected benefits in Table 2 are collated from these sources[74-75].

**B. Barriers to the Achievement of the Expected Benefits**

According to Tidd, et. al. (2005)[76], understanding user needs has always been a critical determinant of innovation success and thus the achievement of expected benefits pertaining to patients Table 1. One way of achieving this is to include users in the software development lifecycle at a much earlier stage, allow them to assume an active role in the innovation process and it will lead to better adoption and higher quality innovation. The user-centred design approach adopted for IT3SS aims to involve the users throughout the entire software development lifecycle (i.e. requirements, design, development, and evaluation phases). According to the EU e-Health Action 2012-2020[77], there are many barriers to the deployment of e-health. These barriers could also impede the achievement of expected benefits listed in Tables 1 and 2. Sustainable innovations in health care practices will not be successful if there is a lack of holistic approach which addresses the interdependencies amongst the following dimensions of the system: users, organisation, technology, standards, legal and regulations, etc...[78]. Thus, in IT3SS, the barriers related to these dimensions will be examined and mitigation measures appropriately taken into order to ensure the achievement of its expected benefits.

**VIII. CONCLUSION**

Porter relates innovation capacity[79] to driving productivity (it will be productivity of technology-enabled health care and services) growth into the future. Innovation systems

thinking[80] emphasises the continuous innovation capacity building of stakeholders rather than a one-off invention. IT3SS will contribute to innovation capacity and integration of new knowledge to meet user needs through the following (adapted from Porter[81]): (i) absorptive capacity building which is a product of learning (measured the ability to acquire, assimilate, transform and exploit knowledge generated from high quality scientific research conducted for the design, development and evaluation of IT3SS system and services)[82]; (ii) provision of a fertile innovation environment via training, technology transfer and knowledge sharing activities; (iii) demanding user need for more efficient and effective products (note: the iterative requirements, design, development and evaluation cycles will bring about incremental innovation to IT3SS products).

**References**

[1] [http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1587763-08\\_health\\_wp2014-2015\\_en.pdf](http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1587763-08_health_wp2014-2015_en.pdf)  
 [2] <http://www.euro.who.int/en/health-topics/health-policy/health-2020-the-european-policy-for-health-and-well-being/about-health-2020>  
 [3] [http://www.healthfirsteurope.org/uploads/Modules/Newsroom/HFE-EqualityInE-health\\_BROCH\\_LayA\\_V23\\_spreads-2.pdf](http://www.healthfirsteurope.org/uploads/Modules/Newsroom/HFE-EqualityInE-health_BROCH_LayA_V23_spreads-2.pdf)  
 [4] [http://personcentredcare.health.org.uk/sites/default/files/resources/what\\_is\\_co-production.pdf](http://personcentredcare.health.org.uk/sites/default/files/resources/what_is_co-production.pdf)  
 [5] <http://www.being.here.net/page/2531/applying-biomimicry-to-global-issues>  
 [6] <http://biomimicry.org/what-is-biomimicry/>  
 [7] Porter, M.I.E. (1985). Competitive Advantage. Ch. 1, pp 11-15. The Free Press. New York.  
 [8] <http://www.ifm.eng.cam.ac.uk/research/dstools/value-chain/>  
 [9] Porter, M.E., and Teisberg, (2006). Redefining Health Care: Creating Value-based Competition on Results, HBS.  
 [10] <http://www.hbs.edu/faculty/Publication%20Files/09-093.pdf>  
 [11] Porter, M.E., and Teisberg, (2006). Redefining Health Care: Creating Value-based Competition on Results, HBS.  
 [12] Ibid  
 [13] <http://www.telehealthcode.eu/>  
 [14] <http://ec.europa.eu/digital-agenda/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century>  
 [15] <http://www.telehealthcode.eu/>  
 [16] <http://www.dynamichealthsystems.co.uk/>  
 [17] <http://www.dynamichealthsystems.co.uk/patients-speaking>  
 [18] <https://ico.org.uk/media/about-the-ico/policies-and-procedures/1043029/information-governance-strategy-2014-2016.pdf>  
 [19] <http://healthcare.orange.com/eng/discover-e-health>  
 [20] <http://healthcare.orange.com/eng/discover-e-health/all-folders/digital-hospital>  
 [21] <http://healthcare.orange.com/eng/discover-e-health/all-use-cases/how-to-combine-high-quality-care-with-a-simplified-routine>  
 [22] <http://healthcare.orange.com/eng/discover-e-health/all-folders/mHealth>  
 [23] [http://ehealth-strategies.eu/database/documents/sweden\\_countrybrief\\_ehstrategies.pdf](http://ehealth-strategies.eu/database/documents/sweden_countrybrief_ehstrategies.pdf)  
 [24] [http://www.ehealthservices.eu/instance/data/prime\\_product\\_julkaisut/np/embds/15969\\_portfolio\\_of\\_ehealth\\_applications\\_final\\_web.pdf](http://www.ehealthservices.eu/instance/data/prime_product_julkaisut/np/embds/15969_portfolio_of_ehealth_applications_final_web.pdf)  
 [25] <http://www.kith.no/upload/6585/BjornAstad-HelsIT2012-T1A-1345.pdf>  
 [26] [http://estemed.org/wp-content/uploads/2013/08/5578-EADV\\_News\\_n47\\_Korolenko.pdf](http://estemed.org/wp-content/uploads/2013/08/5578-EADV_News_n47_Korolenko.pdf)  
 [27] <http://www.wellnesslayers.com/about-us/overview.aspx>  
 [28] <http://www.wellnesslayers.com/about-us/contacts>  
 [29] <http://www.wellnesslayers.com/solutions/our-approach>  
 [30] <http://www.wellnesslayers.com/solutions/our-formula>  
 [31] <http://www.wellnesslayers.com/solutions/products/custom>  
 [32] <http://cs.ecs.baylor.edu/~maurer/CSI5493C/Papers/JASCo.pdf>  
 [33] <https://www.honeywelllifecare.com/>  
 [34] <https://www.honeywelllifecare.com/lifestream-products/legacy-products-and-software/sentry-telemonitor/>  
 [35] <http://www.mulesoft.com/resources/cloudhub/tpaas-integration-platform-as-a-service>  
 [36] <http://www.itgovernance.co.uk/data-protection.aspx#VRpZQnF-M4>  
 [37] <http://www.itgovernance.co.uk/data-protection-dpa-and-eu-data-protection-regulation.aspx#VRpZQnF-M4>  
 [38] <http://azure.microsoft.com/en-gb/support/trust-center/compliance/>  
 [39] <http://www.iso.org/iso/home/standards/management-standards/iso27001.htm>  
 [40] <https://www.gov.uk/government/collections/cloud-security-guidance>  
 [41] <http://www.h17.org/Implement/standards/>  
 [42] <http://cdn.intechopen.com/pdfs/wm/5912.pdf>  
 [43] <http://telecareaware.com/what-is-telecare/>  
 [44] [http://yhlicc.org.uk/wp-content/uploads/2011/10/11092020\\_tele\\_consultation\\_workbk.pdf](http://yhlicc.org.uk/wp-content/uploads/2011/10/11092020_tele_consultation_workbk.pdf)  
 [45] <http://www.nhs.uk/NHSEngland/AboutNHSservices/Emergencyandgeneticservices/Pages/NHS-111.aspx>  
 [46] <http://www.wedo-partnership.eu/food-practice/1177ae-website-and-helpline-healthcare-services>  
 [47] Linoff, G. S., and Berry, M. J. A. (2011). Data Mining Techniques, Wiley Publishing Inc.  
 [48] <http://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare>  
 [49] Kor, A. L., et al. (2011). Chapter 6: A Survey of Epistemology and Its Implications for an Organisational Information and Knowledge Management Model, In. Earley, A., et al. (eds). Innovative Knowledge Management: Concepts for Organizational Creativity and Collaborative Design, Information Science Reference.  
 [50] <http://bma.org.uk/-/media/Files/PDFs/Practical%20Advice%20at%20work/Ethics/socialmediaguidance.pdf>  
 [51] <http://www.w3.org/WAI/intro/usable>  
 [52] <http://www.ngroup.com/articles/ten-usability-heuristics/>  
 [53] <http://www.designprinciplesftw.com/collections/sneidermans-eight-golden-rules-of-interface-design>  
 [54] <http://www.w3.org/WAI/intro/usable>  
 [55] <http://www.legislation.gov.uk/ukpga/2010/15/contents>  
 [56] [http://www.healthfirsteurope.org/uploads/Modules/Newsroom/HFE-EqualityInE-health\\_BROCH\\_LayA\\_V23\\_spreads-2.pdf](http://www.healthfirsteurope.org/uploads/Modules/Newsroom/HFE-EqualityInE-health_BROCH_LayA_V23_spreads-2.pdf)  
 [57] <http://www.telehealthcode.eu/>  
 [58] Norman, D.A., et al. (1986). User Centered System Design: new Perspectives on Human-Computer Interaction, Hillsdale: Lawrence Erlbaum Associates Publishers.  
 [59] F. E. Ritter et al. (2014). Foundations for Designing User-Centered Systems, Springer-Verlag London, DOI: 10.1007/978-1-4471-5134-0\_2.  
 [60] <https://www.deming.org/theman/theories/pdsacycle>  
 [61] [http://www.healthfirsteurope.org/uploads/Modules/Newsroom/HFE-EqualityInE-health\\_BROCH\\_LayA\\_V23\\_spreads-2.pdf](http://www.healthfirsteurope.org/uploads/Modules/Newsroom/HFE-EqualityInE-health_BROCH_LayA_V23_spreads-2.pdf)  
 [62] <http://ailemethodology.org/>  
 [63] IBM. (2012). Agile For Dummies©, IBM Limited Edition, Published by John Wiley & Sons, Inc.  
 [64] <http://www.usability-bremen.de/wp-content/uploads/2013/03/Winter-Holt-et-al-2012-Persona-driven-agile-development.pdf>  
 [65] <http://www.ie.jiit.edu.in/html/uploads/files/2007101010520670.pdf>  
 [66] Roberts, D., Berry, D., Isemsee, S., Mullaly, J. (1998). Designing for the user with OVID, MacMillan.  
 [67] Tidd, J., et al. (2005). Managing Innovation: Integrating technological, market, organizational change, John Wiley.  
 [68] <https://www.healthvault.com/gb/en>  
 [69] <https://www.healthvault.com/gb/en/overview>  
 [70] <http://www.thehealthlabs.com/>  
 [71] [http://www.ehealth-connection.org/files/conf-materials/Global%20Health%20-%20Measuring%20Outcomes\\_0.pdf](http://www.ehealth-connection.org/files/conf-materials/Global%20Health%20-%20Measuring%20Outcomes_0.pdf)  
 [72] Ibid  
 [73] [https://secure.chi.ca/free\\_products/HI2013\\_EN.pdf](https://secure.chi.ca/free_products/HI2013_EN.pdf)  
 [74] <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3278097/>  
 [75] <http://www.asia.sinica.edu.tw/~ccchiang/GILIS/LIS/p9-Delone.pdf>  
 [76] Tidd, J., et al. (2005). Managing Innovation: Integrating technological, market, organizational change, John Wiley.  
 [77] [http://ec.europa.eu/information\\_society/newsroom/dae/document.cfm?doc\\_id=4188](http://ec.europa.eu/information_society/newsroom/dae/document.cfm?doc_id=4188)  
 [78] <http://www.jmir.org/2011/4/e111/>  
 [79] <http://www.ims-productivity.com/page.cfm/content/Creating-Innovative-Capacity/>  
 [80] <http://www.innogen.ac.uk/downloads/Building-the-Case-for-National-Systems-of-Health-Innovation.pdf>  
 [81] <http://www.ims-productivity.com/page.cfm/content/Creating-Innovative-Capacity/>  
 [82] [http://performance.ey.com/wp-content/uploads/downloads/2012/02/Perfo\\_2-January-2012-Journal-v17-Increasing-innovative-capacity.pdf](http://performance.ey.com/wp-content/uploads/downloads/2012/02/Perfo_2-January-2012-Journal-v17-Increasing-innovative-capacity.pdf)

# A Pilot Study on the Web-based Health Promotion Program Combined with a Wearable Device to Increase Physical Activity

Ya-Ting Yang<sup>1</sup> and Jeen-Shing Wang<sup>2</sup>

<sup>1</sup>Institute of Education, National Cheng Kung University, Tainan 701, Taiwan, R.O.C.

<sup>2</sup>Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan, R.O.C.

**Abstract** - *The purpose of this study is to conduct a web-based health promotion program combined with a wearable activity sensor system to promote adult physical activity. The objectives of this pilot study include: 1) the investigation of the influences of sensor wearing behavior on the participants' health outcomes in a real-life setting, and 2) the examination of the effectiveness of sensor wearing duration and intensity of participation. The results of this study demonstrate: 1) sensor wearing can raise participants' awareness of being physically inactive, and consequently can motivate participants to engage in self-promotion of physical activity; 2) health promotion interventions combining wearable activity sensor system with the Internet real-time feedback program can enhance participants' self-monitoring capability and promote weight loss in sedentary overweight adults; and 3) participants who wear the sensor for a longer duration and have a higher level of participation obtain significant improvement in participants' anthropometric and physical fitness values.*

**Keywords:** Worksite health intervention, accelerometer, physical activity, web-based intervention.

## 1 Introduction

The lack of a physically active lifestyle produces negative consequences for sedentary individuals. They are at greater risk for many chronic diseases, often incurring the loss of function and mobility while also having higher mortality rates [1]. Physical activity may help reduce the risk of cardiovascular disease, diabetes, cancer, hypertension, obesity, depression, and osteoporosis while preventing premature death [2]. In Taiwan, only 24.8% of adults (aged from 18 to 65 years) are physically active, with approximately 33% engaging in moderate levels of physical activity, with the remaining 43% engaging in low levels of aerobic activity (defined as being physically inactive) [3]. Since the majority of Taiwanese adults are not (enough) physically active, a high prevalence of adult obesity (42.2% of adults were either obese or overweight) was found in a nationwide, cross-sectional population-based survey study [4].

To reduce the number of obese adults, effective behavior change programs need to be implemented in order to increase adult physical activity. To promote physical activity, helping individuals to become aware of the fact that they are not physically active is important.

Research has discovered that awareness can help to promote physical activity as individuals perceive the risks of their behavior [5]. Electronic wearable devices that provided objective measurement of physical activity and offered real-time feedback did improve individuals' self-monitoring of their physical activity levels. Consequently, their amount of weight loss increased [6]. However, few studies were conducted based on Asian population. Therefore, the purpose of this study was to conduct a web-based worksite health intervention combined with a wearable sensor to promote adult physical activity. The objectives of this pilot study were to: 1) investigate the influences of sensor wearing behavior on the participants' health outcomes in a real-life setting, and 2) examine the effectiveness of sensor wearing duration and intensity of participation.

The rest of this study is organized as follows. In Section 2, we introduce the wearable system and the physical activity estimation algorithm. The MET mapping regression model construction are presented in Section 3. Then, in Section 4, the experimental results effectively validated the proposed approach. Finally, the conclusions are given in the last section.

## 2 Method

### 2.1 Participants

Participants were employees from a telecom company in Taiwan. They were recruited during the annual health examination that the company held for its employees. The study was approved by the Ethical Committee for Human Research at the National Cheng Kung University Hospital. At baseline, a total of 122 employees were interested and volunteered to participate in the project. All participants were office workers and had regular access to the internet. All participants gave their written informed consent. However, 24 participants did not finish the study.

Consequently, only 98 participants' data (76 males and 22 females with an average age of 52 years old) were analyzed. The basic demographic information such as education level and marital information of the dropouts was not significantly different from the participants who completed the study. For the working style, 62.2% of the participants had a sedentary working style while 36.7% of the participants had an active working style.

## 2.2 Study Design and Procedure

The present study had a prospective observational design. Participants' anthropometric information (Body Mass Index, BMI, and waist circumference) and physical fitness ability (cardiorespiratory endurance, muscle strength, muscle endurance, and flexibility) were measured at baseline (before they wore the sensor) and at the one-month follow up (immediately after they stopped wearing the sensor). Participants' physical fitness was measured in order to observe their ability to perform daily tasks. Physical fitness contains two categories: health-related fitness and skill-related fitness. The health-related fitness, including cardiorespiratory endurance, muscular endurance, muscle strength, and flexibility, was tested because this is more important to public health than athletic ability. In this study, physical activity was measured objectively by a wearable activity sensor, while participants' subjective physical activities were measured by a self-administered questionnaire, the International Physical Activity Questionnaire (IPAQ). In addition, the self-administered and short version of the IPAQ questionnaire was used to collect participants' moderate and vigorous physical activity levels for the previous seven days. Based on their BMI score, participants were classified as: 1) underweight when BMI < 18.5 kg/m<sup>2</sup>, 2) normal weight when BMI is between 18.5 to 22.9 kg/m<sup>2</sup>, 3) at risk of obesity when BMI is between 23.0-24.9 kg/m<sup>2</sup>, 4) obesity I (BMI 25.0-29.9 kg/m<sup>2</sup>), and 5) obesity II (BMI > 30.0 kg/m<sup>2</sup>). Central obesity is defined as 1) when men's waist circumference is > 90 cm, and 2) when women's waist circumference is > 80 cm.

Each participant was given three sensors to wear on their wrist, waist, and ankle to record both their upper and lower body movements. The participants were asked to wear the sensor at any time or place during the assessment period, and to upload the recorded signals to the web program of a cloud server system. The activity sensor should avoid contact with water. Based on the uploaded data, the server system automatically analyzed the data and provided individualized health information including steps taken, distance traveled, calories burned, and activity intensity levels. Each participant was authorized to login his/her web program account to review the analysis results categorized as daily, weekly, and/or monthly personal health information. It can be displayed in either Chinese or English. In addition to the analysis results, immediate feedback regarding whether participants' physical activity levels met the recommended

guidelines were provided. The web program offers suggestions on how to improve physical activity levels based on participants' age and weight. A toll free phone number and Internet Q&A forum were available to answer any questions regarding the activity sensor, such as uploading the data, and/or questions related to the suggested intervention.

## 2.3 Instrument

Procedures for the four physical fitness tests were based on the guidelines provided by the Bureau of Health Promotion, Department of Health R. O. C. (Taiwan) [7]. Researchers supervising the physical fitness tests were trained by a professor from the Department of Nursing at National Cheng Kung University. The scores recorded from IPAQ can be quantified into Metabolic Equivalent Task (MET). MET 3-6 represents moderate intensity of physical activity, while MET >6 represents vigorous intensity. One MET is equal to the energy cost during rest and is roughly equal to 3.5ml O<sub>2</sub> kg<sup>-1</sup>min<sup>-1</sup> in adults. The validity analysis indicated that the IPAQ questionnaire has acceptable validity in healthy adults [8]. The activity sensor, a wearable accelerometer, utilized for recording participants' objective physical activity, was developed by the Computational Intelligence & Learning Systems (CILS) Research Lab, located in the Department of Electrical Engineering at National Cheng Kung University. The sensor is a wearable system with high data compression and low power consumption and was developed for recording long-term acceleration for daily physical activity. Based on acceleration signals, automatic analysis algorithms for physical activity measures were developed. The effectiveness of the wearable sensor system and algorithms has been successfully validated by applications for physical activity measures.

## 2.4 Statistical analysis

A paired-sample *t* test was calculated to compare the differences between baseline data (before participants wore the activity sensor) and data from the follow-up test (immediately after participants stopped wearing the activity sensor). Data were analyzed using SPSS version 17. Although participants were told to wear the sensor for a total of one month, some participants did not wear the sensors every day during the test period. In this study, in order to test the influence of sensor wearing duration on participants' health outcomes, two categories of sensor wearing days were compared: 1) differences between the sensor wearing days < 5 days and > 5 days, and 2) differences between the sensor wearing days < 10 days and > 10 days. Additionally, to investigate the influence of participation intensity on participants' health outcomes, the number of files uploaded for analysis was selected. Participants who uploaded less than or equal to 8 data files were classified as at a lower level of participation. Participants who uploaded more than or equal to 9 data files were classified as at a higher level of

Table 1 Differences in the duration of sensor wearing days more than 10 days or less in health outcomes

Variable		Sensor wearing days <10 days n=75		Sensor wearing days >10 days n=23		Between group
		mean±SD	<i>p</i> <sup>a</sup>	mean±SD	<i>p</i> <sup>a</sup>	<i>p</i> <sup>b</sup>
BMI	Baseline	26.2±3.5	NS	24.5±3.1	NS	< .05
	Follow-up	26.2±3.4		24.3±2.9		
Sit and reach (cm)	Baseline	26.9±13.5	NS	24.0±12.1	< .01	NS
	Follow-up	28.2±12.2		25.8±12.1		
1 minute sit ups	Baseline	17.2±10.0	< .01	16.4±9.8	< .01	NS
	Follow-up	19.7±9.7		19.6±10.7		
Waist circumference	Baseline	89.1±8.6	< .05	87.7±9.1	< .05	NS
	Follow-up	86.9±9.9		85.0±9.		
Hand grip(Kg)	Baseline	38.4±8.9	< .01	40.3±10.6	< .01	NS
	Follow-up	39.8±9.2		42.1±10.7		
3 minutes of walking up stairs	Baseline	61.6±10.0	< .01	57.2±8.6	< .05	NS
	Follow-up	67.2±14.2		62.2±8.3		
Physical activity (MET/hour/wk)	Baseline	2334.2±2588. 2	NS	1657.8±1788.5	< .01	NS
	Follow-up	2845.7±2545. 3		3234.8±2052.9		

<sup>a</sup> *p* value was examined by paired-t test.

<sup>b</sup> the between groups were examined by ANCOVA for adjusting the confounder from baseline.

participation. The paired-t sample tests were performed to analyze the differences between the baseline and follow-up data for sensor wearing duration and intensity of participation. The between group analyses for sensor wearing duration and for intensity of participation were conducted using ANCOVA.

### 3 Results

Participants' anthropometric values indicated that 48.4% of the participants were obese or overweight based on BMI. Among them, 7 participants (5.7%) were severely obese (BMI greater than 30). Based on the waist circumference result, the number of abdominally obese or overweight participants was 51.64%. Significant differences were detected for waist circumference ( $p < 0.01$ ), one minute of sit up ( $p < 0.001$ ), handgrip ( $p < 0.001$ ), 3 minutes of walking up stairs ( $p < 0.001$ ), and physical activity ( $p < 0.05$ ). After one month of wearing the activity sensor, participants had a

significantly lower value of waist circumference and scored significantly higher in 1 minute of sit up, handgrip, and 3 minutes of the walking upstairs test, as well as self-reported physical activity.

To investigate whether the significant differences were due to sensor wearing, a Pearson's Correlation Coefficient was conducted. Based on the statistical results, the number of files uploaded (0.28) and the duration of sensor wearing (0.32) showed a strong positive correlation with physical activity. Significant positive correlations were shown between flexibility and the number of files uploaded (0.22) and the duration of the sensor wearing (0.22). Physical activity and BMI had a strong negative correlation (-0.27). The statistical analysis found no significant differences between the sensor wearing days < 5 days and the sensor wearing duration > 5 days for participants' health outcomes. However, significant differences started to show for participants' health outcomes when the sensor wearing duration was greater than 10 days.

Based on the results in Table 1, participants who wore the sensors for a shorter duration (<10 days) had significant improvement in waist circumference ( $p<0.05$ ), 1 minute sit ups ( $p<0.01$ ), handgrip ( $p<0.01$ ), and 3 minutes of walking up stairs ( $p<0.01$ ) at follow-up compared to the baseline data. Participants who wore the sensors for a longer duration (>10 days) showed significant improvement in sit and reach ( $p<0.01$ ), 1 minute sit up ( $p<0.01$ ), waist circumference ( $p<0.05$ ), handgrip ( $p<0.01$ ), 3 minutes of walking up stairs ( $p<0.05$ ), and physical activity ( $p<0.01$ ) at follow-up compared to at baseline. Results suggest that participants who wore the sensors (> 10 days) for a longer duration had significantly lower BMI scores ( $p<0.05$ ) compared to participants who wore the sensors for a shorter duration (<10 days).

Our results show that participants who had a lower level of participation had significant improvement in handgrip ( $p<0.001$ ) and 3 minutes of walking up stairs ( $p<0.01$ ) at follow-up compared to at baseline. Participants who had a higher level of participation obtained significant improvement in sit and reach ( $p<0.001$ ), 1 minute sit up ( $p<0.001$ ), waist circumference ( $p<0.01$ ), handgrip ( $p<0.001$ ), 3 minutes of walking up stairs ( $p<0.001$ ), and subjective physical activity ( $p<0.01$ ) at follow-up compared to at baseline. Based on our statistical results, participants with more intense participation (> 9 uploaded data files) had significantly lower BMI values as well as higher sit and reach scores compared to participants with less intense participation (< 8 uploaded data files). Participants' total calorie consumption was calculated based on the objectively measured physical activities from the sensors. Statistical analysis indicated that participants' calorie consumption was significantly higher at follow-up compared to at baseline ( $t=2.07$ ,  $p<0.05$ ).

#### 4 Discussion

Results from the present study indicated that sensor wearing behavior would encourage participants to become more active as their physical fitness ability, including cardiorespiratory endurance, muscle strength, and muscle endurance improved significantly after one month of sensor wearing. These significant findings may be attributed to the elevated awareness of participants after sensor wearing of being physically inactive. Immediately, sensor wearing could raise participants' awareness of being physically inactive, and would consequently cause participants to engage in self promotion of physical activity.

The present study indicated that participants' waist circumference values significantly decreased after one month of sensor wearing. The sensor could be an effective device for self-monitoring of weight loss. In addition to electronic devices that promote self-monitoring, the current study also offered the benefits of real-time feedback for behavior modification through the Internet. In this study, the duration for wearing the sensor and the intensity of participation had

significant impact on participants' anthropometric and physical fitness values.

There are several limitations in this study. First, participants were volunteers and not randomly selected from the general population, which could limit its generalization. Second, the current study was intended to be implemented in a real-life setting where the sensor and the web program could be introduced to the participants to promote physical activity. However, the limitation of this approach was the lack of a control group. Third, this pilot research was conducted in a small sample size and a short program period. Effects on improving body composition and physical fitness may not be able to fully show the effectiveness of the sensor system and health promotion program in such a short period of time and small sample size. These factors associated with the limitations in this pilot study will be addressed in our future study.

#### 5 References

- [1] Haskell WL, Blair SN, Hill JO. Physical activity: health outcomes and importance for public health policy. *Preventive Medicine* 2009;49:280-2.
- [2] Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. *Canadian Medical Association Journal* 2006;174:801-09.
- [3] Bauman A, Bull F, Chet T, Craig CL, Ainsworth BE, Sallis JF, et al. The international prevalence study on physical activity: results from 20 countries. *International Journal of Behavioral Nutrition and Physical Activity* 2009;6:21.
- [4] Hwang LC, Bai CH, Chen CJ. Prevalence of obesity and metabolic syndrome in Taiwan. *Journal of the Formosan Medical Association* 2006;105:626-35.
- [5] Ronda G, Van Assema P, Brug J. Stages of change, psychological factors and awareness of physical activity levels in the Netherlands. *Health Promotion International* 2001;16(4):305-314.
- [6] Shuger SL, Barry VW, Sui X, McClain A, Hand GA, Wilcox A, et al. Electronic feedback in a diet- and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity* 2011;8:41.
- [7] Bureau of Health Promotion, Department of Health, R.O.C. (Taiwan), retrieved November 29, 2011 from [http://obesity.bhp.gov.tw/cht/index.php?code=list&flag=detail&ids=129&article\\_id=288](http://obesity.bhp.gov.tw/cht/index.php?code=list&flag=detail&ids=129&article_id=288)
- [8] Hagströmer M, Oja P, Sjöström M. The international physical activity questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutrition* 2006;9(6):755-62.

# Immediate Lead Positioning Feedback for ML based wearable ECG

A. Rahul Krishnan<sup>1</sup>, B. Maneesha Vinodini Ramesh<sup>1</sup>

<sup>1</sup>Amrita Center for Wireless Networks and Applications, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India

**Abstract**—Wearable ECG devices are used in combination with smartphones to capture, analyze and transmit ECG data to far away hospitals where doctors analyze, diagnose and provide necessary patient feedback. ECG electrode cable misplacement is a significant challenge when untrained patients use them, especially among rural population. Lead misplacement algorithms exist for standard ECG electrode positioning, but poses different challenges for Mason-Likar placement, which is most suited for ambulatory ECG measurements. We discuss ECG morphology based lead placement detection algorithms for Mason-Likar configuration, that can provide immediate feedback to patients.

**Keywords:** Smartphone ECG Analysis, Mason-Likar lead misplacement

## 1. Introduction

Wearable ECG devices help in health monitoring in remote areas. ECG measurements are usually taken by primary health workers (PHW) or patients themselves and sent for professional diagnosis to doctors in hospitals. About 2-3% of all the ECG measurements taken in clinics have wrong measurements due to misplacement or other human errors as shown in [1].

Such unassisted electrode placements can lead to misplacements and generate valid, but false positive disease morphologies. Various techniques have been proposed to detect lead reversals [2], [1], [3] for resting ECG measurements. In the case of wearable monitoring, the placement of leads are made according to Mason-Likar, which allows the patients to continue his activities even while the ECG is monitored. The change of placements in Mason-Likar (ML) introduces different characteristics in the ECG morphology and along with it challenges in detecting any misplacement. Using existing algorithms, these lead misplacements can be detected and analyzed for certain features that can later be used for real-time feedback.

## 2. Related Work

ECG lead misplacement is a widely studied subject, both medically and in terms of detection algorithms. Various techniques to detect and differentiate them, both manual and automatic exist. Batchvarov et al. [4] has analyzed all the possible electrode misplacement and their consequence on the resulting ECG morphology. Techniques suggested in [5], [1], [2], [3] have been used to automate the detection

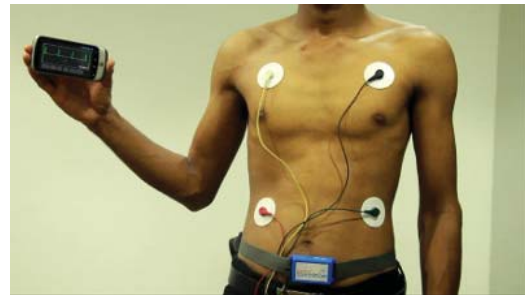


Fig. 1: Wearable ECG device along with the 4 electrode cables connected according to Mason-Likar placement. The smartphone shows the live ECG data.

of lead misplacement. With the larger adoption of wearable ECG devices, the requirement for real-time analysis and feedback has opened up many challenges. Most of the work is based on standard ECG lead placements, whereas in case of wearable ECG devices, the electrodes are placed according to Mason-Likar lead configuration. A detailed study on the effect of ML lead misplacement is unavailable in the current literature.

Han et al. [6] proposed an algorithm that can detect lead misplacement accurately in both standard and ML ECG signals. But misplacement involving the right leg electrode was not considered in their study. A patient feedback system that can provide real-time feedback to patients based on changes in lead positioning is not yet available and our work in this area aims to fill this research gap.

## 3. System

Electrode cables are placed according to the ML lead placements, i.e., arm electrodes (RA and LA) are placed at the infraclavicular fossae and the leg electrodes (RL and LL) on the lower abdomen. This allows the patients to move around and engage in daily activities. Fig 1 shows the ECG device and the 4 electrodes placed on the torso of a patient. The wearable is also connected to a mobile device which runs an Android app. This is used to record ECG data from the patient for later analysis.

Three electrode misplacements that we are studying are: LA-LL, RA-LL and LA-RA.

## 4. ML Vs. Standard ECG

Mason-Likar system changes the morphologies that are used to detect lead misplacement in standard configuration.

Table 1: Comparison of ECG morphologies between standard and ML electrode placements.

Placement	RA-RL	LA-RL	Dual
Standard ECG	$II = 0, -I$	$III = 0, aVR \approx -II$	$I = 0, -III, aVF \approx III$
ML ECG	$-I$	$aVR \approx -II$	$-III, aVF \approx III$

Table 2: QRS Axis Measurement for normal and misplaced electrodes.

QRS Axis	Normal	LA-LL	RA-LL	LA-RA
Avg	75.1	24.1	147.6	105.5
( $\pm Sn * t$ )	( $\pm 1.4$ )	( $\pm 5.5$ )	( $\pm 5.1$ )	( $\pm 2.2$ )

Lead I amplitude is reduced while Lead II amplitude increases. Table 1 describes these differences.

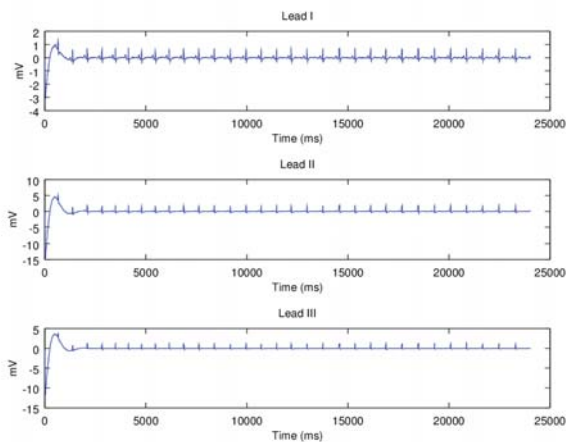


Fig. 2: Leads I, II and III when LA-RL are interchanged.

## 5. Method

### 5.1 Data Collection

We obtained the ECG data from MIT-BIH database and smartphone recordings of healthy subjects in both correct and lead misplaced scenario for about 20s. The stored data was later analyzed using Octave. The average of ten R peak amplitudes of lead I and aVF was used to calculate the QRS axis. This data was used to analyze QRS axis based algorithms for its ability to differentiate between lead misplacements.

### 5.2 Analysis

Fig 3 shows a box plot of the QRS axis measurement data. The bars represent the standard error of means (SEM) with  $p=0.05$  (Average  $\pm Sn * t$ ). The ECG signals with correct electrode placements have QRS axis of 75.079 ( $\pm 1.38$ ). It may be noted that there is significant difference between LA-LL, RA-LL, LA-RA. Therefore, these three misplacements

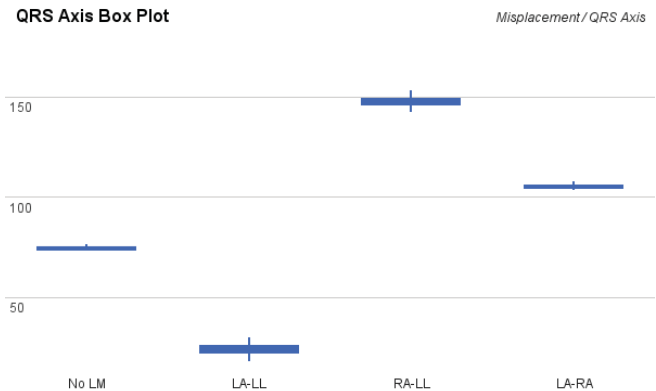


Fig. 3: QRS axis measurement for normal and 3 electrode misplacements.

can be differentiated between themselves and also from the normal placement. RA-LL and LA-RA showed significant right axis deviation too.

## 6. Algorithm

Based on these results, a QRS axis based classification algorithm would be able to differentiate the three above said misplacements. To achieve real-time analysis of ECG signal, we will need to use initial filtering of data and consequently calculate the QRS axis measurement. The filtered data consists of lead I and lead II data. Other four leads (III, aVR, aVL and aVF) are then derived from the available data. The average of ten R peaks each in lead I and aVF is used to calculate QRS axis. Using QRS axis, we classify the lead misplacement into four categories: Normal (75.1 ( $\pm 1.4$ )), LA-LL (24.1 ( $\pm 5.5$ )), RA-LL (147.6 ( $\pm 5.1$ )), LA-RA (105.5 ( $\pm 2.2$ )).

## 7. Conclusion

It has been shown that QRS axis can be used to detect and differentiate lead misplacements in Mason-Likar based ECG data. The results are encouraging and we expect to conduct a more extensive study of various possible lead misplacements. The analysis using Octave shows that lead misplacement detection system can classify 3 different lead misplacements in ML configuration.

## 8. Future Work

The area of lead positioning feedback needs further research especially in terms of extensively studying all the

possible misplacements and exploring other measures of ECG signals apart from QRS axis. An algorithm that can detect misplacements in real-time, coupled with the wearable ECG device can reduce the need for technicians and doctors to provide feedback on the validity and quality of the captured ECG data.

## 9. Acknowledgement

We would like to extend our deepest gratitude to our Chancellor Sri Mata Amritanandamayi Devi for guiding, inspiring and supporting us in carrying out research for betterment of healthcare delivery among rural population. We also acknowledge the contribution of volunteers who supported us in the data collection process.

## References

- [1] B. Hede, M. Ohlsson, L. Edenbrandt, R. Rittner, O. Pahlm, C. Peterson, *et al.*, "Artificial neural networks for recognition of electrocardiographic lead reversal," *The American journal of cardiology*, vol. 75, no. 14, pp. 929–933, 1995.
- [2] J. A. Kors and G. van Herpen, "Accurate automatic detection of electrode interchange in the electrocardiogram," *The American journal of cardiology*, vol. 88, no. 4, pp. 396–399, 2001.
- [3] V. Chudáček, L. Zach, J. Kuzilek, J. Spilka, and L. Lhotská, "Simple scoring system for eeg quality assessment on android platform," in *Computing in Cardiology, 2011*. IEEE, 2011, pp. 449–451.
- [4] V. N. Batchvarov, M. Malik, and A. J. Camm, "Incorrect electrode cable connection during electrocardiographic recording," *Europace*, vol. 9, no. 11, pp. 1081–1090, 2007.
- [5] I. Jekova, V. Krasteva, I. Dotsinsky, I. Christov, and R. Abacherli, "Recognition of diagnostically useful eeg recordings: Alert for corrupted or interchanged leads," in *Computing in Cardiology, 2011*. IEEE, 2011, pp. 429–432.
- [6] C. Han, R. E. Gregg, and S. Babaeizadeh, "Automatic detection of eeg lead-wire interchange for conventional and mason-likar lead systems," *interchange*, vol. 8, p. 9.



# An Energy Expenditure Estimation Algorithm for a Wearable System

Jeen-Shing Wang<sup>1</sup>, Fang-Chen Chuang<sup>1</sup>, and Ya-Ting Yang<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan, R.O.C.

<sup>2</sup>Institute of Education, National Cheng Kung University, Tainan 701, Taiwan, R.O.C.

**Abstract** - This paper presents of an algorithm for physical activity classification and MET mapping regression model construction for a wide range of daily activities using a wearable system. The sensor system consists of several sensor modules that can be synchronized to record the accelerations of diverse motions/activities. During the measurement the accelerations of daily activities, three sensor modules are worn at the participant's hand wrist, waist, and ankle, respectively. In addition, the participant's chest is attached to an indirect calorimeter (Cosmed K4b<sup>2</sup>) to measure oxygen uptake to calculate actual metabolic equivalent (MET) during the experiments. The oxygen uptake for different activities is used to construct MET mapping regression models. Our experimental results shows that the average classification accuracy of five categories of physical activities is 95.33%. The average error of MET estimation without and with activity classification is  $-7.25 \times 10^{-15} \pm 1.16$  METs and  $2.31 \times 10^{-4} \pm 0.71$  METs, respectively.

**Keywords:** Accelerometer, metabolic equivalent, neural networks, energy expenditure.

## 1 Introduction

Nowadays, many people suffer from physical inactivity due to lifestyle changes. This phenomenon not only increases the incidence of chronic illness to cause a burden of medical resources but also seriously affects living quality. Much literature [1], [2] has pointed out the importance of performing physical activities to prevent chronic illness. For example, having enough physical activities could reduce the incidence of certain diseases, such as coronary heart disease, apoplexy, high blood pressure, diabetes, breast cancer, etc. Hence, in the past years, it has been an important issue that how to help people find out whether there is enough activity and energy expenditure (EE) in their daily lives. The existing tools for evaluating energy expenditure of physical activities include activity questionnaires [3], indirect calorimeter [3], heavy water [3], [4], electrocardiograph [5], [6], and accelerometers [7], [8]. However, when using any of the tools to estimate physical activities in daily life, we have to consider certain characteristics, such as accuracy, convenience, low cost, high comfort, portable, and automatically monitoring and recording. Therefore, accelerometers have been utilized in the research field to

characterize the intensity and duration of physical activity, and their output has been used for the estimation of energy expenditure [9], [10], [11].

Accelerometer devices are typically attached to wrists [6], [12], arms [13], waists [12], [14], hips [15], thighs [6], or ankles [12] respectively to estimate the intensity of physical activity and energy expenditure. A common EE estimation method is to utilize the acceleration counts generated by body motions from an accelerometer, personal parameters (height and weight), and the actual metabolic equivalent (MET) measured by the Cosmed K4b2 to construct a single linear MET mapping regression model [16], [17]. The advantages of the EE estimation methods are low computational complexity and easy feasibility, such as Actigraph, Actical, AMP-331 and Actiwatch. However, using a single linear regression model cannot accurately estimate EE for a wide range of physical activities. For example, Crouter *et al.* [9] compared the accuracy of the regression models developed by Actigraph, Actical and AMP-331, respectively.

Actigraph and Actical overestimated the energy cost of walking and static activities, while seriously underestimated that of vigorous activities. AMP-331 gave a close estimation of the energy cost during walking, but slightly overestimated sedentary/light activities and underestimated vigorous activities. Therefore, in order to improve the accuracy of energy expenditure estimation, there are multiple regression models developed after activity classification [9], [18]. These researchers asked the participants to wear an accelerometer on his/her limb or trunk and used one or multiple threshold values to divide an extensive activity into several sectors, developing linear or nonlinear models for each sector to improve estimation accuracy. To name a few, Crouter *et al.* [18] wore a uniaxial Actical at waist. When count value  $\leq 10$ , they set MET = 1, and when count value  $> 10$ , the coefficient of variation (CV) would be set as 13 to be the threshold value. If  $CV \leq 13$ , an exponential model is used to estimate walking and running activities, and the coefficient of determination (R<sup>2</sup>) and the standard error of the estimation (SEE) are 0.912 and 0.149 METs, respectively. Otherwise, a lifestyle/leisure time physical activity regression model was used with a  $CV > 13$ , and  $R^2 = 0.884$  and  $SEE = 0.804$ . Klippel and Heil [9] used Actical mounted on hip to calculate counts, and further used different threshold count values to differentiate between four activity intensity levels, which were 1)  $< 50$ , 2)  $50 \leq$  counts  $< 350$ , 3)  $350 \leq$  counts  $< 1200$ , and 4)  $\geq 1200$ . When the count value fell within level 1 and 2, the resulting METs were

set as 0.9 METs and 1.83 METs, respectively. When the count value fell within level 3 and 4, two regression models were developed to estimate METs for level 3 and 4, respectively. The estimated results of level 3 and 4 were  $R^2 = 0.74$  and  $SEE = 0.8$  METs, and  $R^2 = 0.84$  and  $SEE = 0.9$  METs, respectively.

In this paper, a wearable system, an activity classification algorithm, and multiple MET mapping regression model were developed. The system consists of several sensor modules that can be synchronized to record motions/activities for physical activity classification with wide range of activity intensities. The proposed physical activity classification algorithm is composed of the procedures of acceleration acquisition, signal preprocessing, and feature generation. The acceleration signals of body motions are measured by the wearable sensor modules. The features extracted from the acceleration signals include count, coefficient of variation, and the ratio of frequency's amplitudes. The features are sent to a probability neural network (PNN) classifier for activity classification. After using the PNN to classify five activity categories, the three count values and personal parameters, height and weight, are utilized to construct five MET mapping regression models to estimate MET for the five activity categories, respectively. The rest of this study is organized as follows. In Section 2, we introduce the wearable system and the physical activity estimation algorithm. The MET mapping regression model construction are presented in Section 3. Then, in Section 4, the experimental results effectively validated the proposed approach. Finally, the conclusions are given in the last section.

## 2 Wearable System and Its Energy Expenditure Estimation Algorithm

In this study, the wearable system developed by our research lab consists of a user-specified number of sensor modules that can be synchronized by computer application software. The module consists of the following major components: an accelerometer (LIS3L02AQ3), a microcontroller (STM32F103C8), a micro SD card, and a battery. The LIS3L02AQ3 possesses a user selectable full scale of  $\pm 2g$  and  $\pm 6g$ , and is able to measure accelerations over the bandwidth of 1.5 KHz for all axes. The accelerometer's sensitivity is set to  $\pm 6g$  in this study. The microcontroller collects the analog signals generated from the accelerometer and transforms the signals to digital ones via an internal 12 bits A/D converter. The digital sampling rate (fs) of the microcontroller is 10 Hz. The overall power consumption of the hardware device is 25~28 mA at 4.2 V. In our experiment, we recruited 13 students between 20 to 30 years old (9 males and 4 females, and their average body mass index (BMI)  $22.37 \pm 2.79$  kg/m<sup>2</sup>), and each participant were asked to wear the wearable system to perform the predefined activities. The subjects were nonsmokers and were free of both diseases and medications known to change their metabolic rates. The wearable system includes three

accelerometer-based portable activity recorders and indirect calorimeter used to collect accelerations and oxygen consumption, respectively.



Fig. 1. A participant wears K4b<sup>2</sup> and the wearable physical activity sensor system. (a) The front view of the participant. (b) The back view of the participant.

### 2.1 Indirect Calorimeter

Each participant wore a portable indirect calorimeter system (Cosmed K4b<sup>2</sup>, Rome, Italy) as Fig. 1 for the duration of performing each activity and resting time. The Cosmed K4b<sup>2</sup> which weighs 1.5 kg, including the battery and a specially designed harness was worn on participants' chest by a chest harness. Also, a flexible face mask was placed over the participants' mouth and nose using a nylon mesh hairnet and Velcro straps to secure. This mask was attached to a flowmeter which is a bidirectional digital turbine and an optoelectronic reader. A disposable gel seal placed between the participant and the face mask was used to prevent air leaks from the face mask. According to the manufacturer's guidelines, the Cosmed K4b<sup>2</sup> oxygen analyzer and carbon dioxide analyzer were calibrated before performing each test. The calibration process had four steps: room air calibration, reference gas calibration, delay calibration, and turbine calibration. First, the room air calibration was automatically run to update the CO<sub>2</sub> analyzer baseline and the O<sub>2</sub> analyzer gain so that they coincided with atmospheric values. We then performed a reference gas calibration using 15.93% oxygen and 4.92% carbon dioxide. After that, the delay calibration was performed in order to adjust for the lag time between the expiratory flow measurement and the gas analysis. The final step was the turbine calibration that set the flowmeter with a 3.00-L syringe (Hans-Rudolph) to guarantee accurate volume measurements. After completing the calibration process, we entered the ambient humidity determined by a hygrometer into the Cosmed K4b<sup>2</sup>. In addition, we also entered participants' physical characteristics (age, height, weight, and gender) into the Cosmed K4b<sup>2</sup>. After finishing the test, we downloaded all data stored in the memory of the portable Cosmed K4b<sup>2</sup> to a PC.

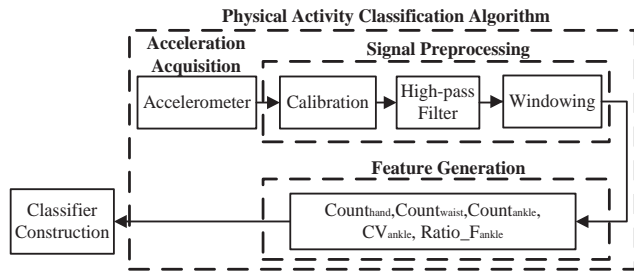


Fig. 2. The block diagram of physical activity classification procedure.

## 2.2 Experimental Procedures

All participants were asked to complete fourteen activities including various lifestyle activities and conditioning exercises that were divided into five categories:

- 1) Category 1 (static): lying quietly, standing quietly, sitting quietly, and doing computer work by sitting.
- 2) Category 2 (home activities): sweeping and mopping.
- 3) Category 3 (walking stairs): walking upstairs and downstairs.
- 4) Category 4 (ambulation): walking at 3 and 4 mph, and running at 6 and 7 mph on a treadmill (SportArt 6310), respectively.
- 5) Category 5 (bicycling): riding an indoor bicycle trainer (SportArt C5150) at 50 and 100 watts.

Before performing the required activities, each participant was instructed to wear three sensor modules as Fig. 1 fixed on the hand wrist, the waist, and the ankle, respectively. The participants all performed each activity for 6 min, and took a rest between activities for at least 5 min to ensure their heart rates were below 100 bpm. We only extracted acceleration signals from the 3<sup>th</sup> to the 6<sup>th</sup> min for analysis. We segmented the acceleration data into non-overlapping windows of 1 min in length and extracted important features from each window.

When each participant performed each activity, oxygen consumption and accelerations were measured continuously throughout the routine by a Cosmed K4b<sup>2</sup> and the sensor modules, respectively. We only extracted oxygen consumption and acceleration signals from the 3<sup>th</sup> to the 6<sup>th</sup> minute for analysis. This time duration was segmented into four periods and each period was equivalent to 1 min. Oxygen consumption was averaged over a 1-min period. Oxygen consumption was converted to METs by 1 MET = 3.5 ml/min/kg = 1 kcal/hr/kg. After introducing the experimental protocol, we used a probability neural network (PNN) [19] to classify five categories and developed a MET mapping regression model for each category to estimate the physical activity's intensity. According to the MET mapping regression model, we could calculate each subject's energy expenditure (unit: kcal), which is each subject's weight (unit: kg) multiplied by METs and the time spent (unit: hr) in the activity.

## 2.3 Physical Activity Classification Algorithm

The block diagram of the proposed physical activity classification algorithm consisting of acceleration acquisition,

signal preprocessing, and feature generation is shown in Fig. 2. In this study, the activities for classification include five categories: static, home activities, walking stairs, ambulation, and bicycling. The acceleration signals of these activities are measured by three sensor modules, and then preprocessed by calibration and filtering. Consequently, the features are extracted from the windowed signal to represent the characteristics of different activities. The characteristics of different body movement signals can be obtained by extracting features from the preprocessed *x*-, *y*-, *z*-axis accelerations. When the procedure of feature generation is done, 5 features are then generated. The features include count [6], coefficient of variation (CV) [7], and the ratio of frequency's amplitudes [8]. A probabilistic neural network (PNN) classifier is used for activity classification. Typically, a PNN consists of an input layer, a pattern layer, a summation layer, and a decision layer. In this paper, the output of the PNN is represented as the label of the desired outcome defined by users. For example, in our physical activity categories, the labels '1', '2', '3', '4', and '5' are used to represent activity categories 1, 2, 3, 4, and 5, respectively. With enough generated features, the PNN is guaranteed to converge to a Bayesian classifier, and thus it has a great potential for making classification decisions accurately and providing probability and reliability measures for each classification. In addition, the training procedure of PNNs only needs one epoch to adjust the weights and biases of the network architecture.

We now summarize the proposed physical activity classification algorithm in the following steps:

**Step 1:** Acquire the raw acceleration signals from three accelerometer-based sensor modules.

**Step 2:** Remove drift errors or offsets by calibration, and remove the gravity from the filtered accelerations by a high-pass filter. Finally, segment the acceleration data into non-overlapping windows of 1 min in length by windowing.

**Step 3:** Generate the time- and frequency-domain features from the preprocessed acceleration of each axis including count, CV, and the ratio of frequency amplitudes.

## 3 Construction of MET Regression Model

We developed MET mapping regression models for energy expenditure estimation. There are two MET mapping regression models developed including a single regression model without activity classification and multiple regression models with activity classification. The parameters of each model were the three counts, subject's weight (WT), and subject's height (HT) as shown in the following equation:

$$MET_e = \sum_{i=1}^n (a_i \times count_{wrist}^i + b_i \times count_{waist}^i + c_i \times count_{ankle}^i) + d \times WT + e \times HT + f, \quad (1)$$

where  $n$  is the power of the model.  $a_i$ ,  $b_i$ ,  $c_i$ ,  $d$ , and  $e$  are coefficients.  $MET_e$  calculated by oxygen consumption is the output of each model. Without activity classification, we utilized the least squares method to develop a single regression model in order to estimate the METs of five

physical activity categories. With activity classification, we still utilized the least square method to develop multiple regression models according to each activity category. For all regression models, we developed from the simple to polynomial mapping regression models. The mean and the standard deviation of estimation errors were utilized to evaluate whether the activity classification is conducive to MET estimation. The mean and the standard deviation of estimation errors can be calculated by the following equations (4) and (5), respectively.

$$SSE = \sum_{i=1}^n (MET_a - MET_e)^2 \quad (2)$$

$$\text{error} = AEE_r - AEE_e, \quad (3)$$

$$\text{error}_{mean} = \frac{\sum_{i=1}^n \text{error}_i}{n}, \quad (4)$$

$$\text{error}_{std} = \sqrt{\frac{\sum_{i=1}^n (\text{error}_i - \text{error}_{mean})^2}{n-1}}, \quad (5)$$

where  $n$  is the total number of the activities conducted by all participants.  $AEE_r$  is the actual MET during activities.  $error$  is the estimation errors of METs.  $error_{mean}$  is the mean of estimation errors.  $error_{std}$  is the standard deviation of estimation errors. The units of  $AEE_r$ ,  $error$ ,  $error_{mean}$ , and  $error_{std}$  are MET.

## 4 Experimental Results

The effectiveness of physical activity classification algorithm is evaluated by the experiments of five activity categories and the results of MET mapping regression models are evaluated by comparisons between without and with classification. The proposed classification algorithm includes acceleration acquisition, signal preprocessing, and feature generation. We employed the PNN as our classifier to classify five categories of physical activities. After physical activity classification, the MET mapping regression model of each activity category is developed to estimate activity energy expenditure. In this study, there were two MET mapping regression models developed including a single MET mapping regression model without classification and multiple MET mapping regression models with classification. The results of energy expenditure estimation of multiple MET mapping regression models with classification will compare with the results of without classification.

### 4.1 Results of Physical Activity Classification

we utilized a leave-one-subject-out cross-validation method to validate the effectiveness of the PNN classifier. The neuron numbers and the computational time of the PNN classifier are shown in Table 1. In each repetition of the cross-validation process, the acceleration data collected from 12 participants randomly was used in the training procedure. Then, the acceleration data from the rest who was left out of the training procedure was used to test the classification performance. The procedure was repeated for all subjects. The

average recognition accuracy was 95.33% in the cross-validation procedure. The best classification accuracy of static activity category was 100%, and the worst classification accuracy of walking stairs was 80.77%. From Table 2, the misclassification of walking stairs category focuses on walking upstairs. The main reason of false alarm is because the acceleration signals of walking upstairs and walking were more similar and thus resulted in more similarity in the extracted features.

Table 1 Neurons and the computational time of the PNN classifier

Numbers of input neuron	5
Numbers of hidden neuron	900
Numbers of output neuron	5
Computational time (sec.)	0.031

### 4.2 Results of MET Estimation

In this study, there are two MET mapping regression models developed including single regression model without activity classification and multiple regression models with activity classification. Table 3 shows the comparisons of the mean and the standard deviation of MET estimation errors by the regression models developed with and without activity classification. Without activity classification, the results of the quartic regression model were the most accurate, where the standard deviation of estimation errors was 1.16 METs. With activity classification, all standard deviations of estimation errors of each power regression model were less than that of the regression models developed without activity classification. Thus, the results validated that the accuracy of MET estimation could be effectively enhanced by using multiple regression models developed with activity classification. The MET estimation errors of the single regression model and the multiple regression models are shown in Fig. 3(a) and (b), respectively.

Without activity classification, the single regression model seriously underestimated the bicycling activity with 2.05 METs less than the actual MET as shown in Table 4. It also overestimated walking stairs activity with 0.67 METs more the actual MET. The results showed that the single regression model could not provide accurate activity MET estimation. On the contrary, constructing multiple regression models supplemented with the activity classifier to estimate METs could decrease the errors of estimation the two activity categories to 0.16 METs and 0.21 METs, respectively, and thus effectively improved the estimation accuracy.

The average AEE estimation values for each category and activity with and without activity classification are shown in Figs. 3 and 4, respectively. Without activity classification, although the single regression model could give a close estimation for light-intensity activities, such as static and home activities, and vigorous-intensity activities, such as running, an estimation error higher than 0.5 METs would occur for six activities including walking upstairs and downstairs, walking (3mph and 4mph), and riding an indoor bicycle trainer (50 watts and 100 watts). Among all, walking upstairs had the most serious estimation error with 1.84 METs in average.

Table 2 Confusion matrix of each activity of physical activity classification for all subjects

Activity categories	Recognized		Static	Home activities	Walking Stairs	Ambulation	Bicycling	Accuracy (%)
	Activities							
Static	Standing		52	0	0	0	0	100.00
	Sitting		52	0	0	0	0	100.00
	Lying		52	0	0	0	0	100.00
	Desk working		52	0	0	0	0	100.00
Home activities	Sweeping		0	52	0	0	0	100.00
	Mopping		1	51	0	0	0	98.08
Walking Stairs	Walking upstairs		0	0	35	17	0	67.31
	Walking downstairs		0	0	49	3	0	94.23
Ambulation	Walking 3mph		0	0	1	51	0	98.08
	Walking 4mph		0	0	0	52	0	100.00
	Running 6mph		0	0	8	44	0	84.62
	Running 7mph		0	0	0	52	0	100.00
Bicycling	bicycling 50 Watts		0	0	0	4	48	92.31
	bicycling 100 Watts		0	0	0	0	52	100.00

Table 3 Means and standard deviations of MET estimation errors by different power of regression models

MET errors of all activities (METs)	Without classification		With classification	
	Mean	Standard deviation	Mean	Standard deviation
	The power of N of the models			
1	$1.11 \times 10^{-15}$	1.49	$-1.58 \times 10^{-16}$	0.8
2	$-3.71 \times 10^{-15}$	1.24	$-2.65 \times 10^{-17}$	0.77
3	$2.95 \times 10^{-15}$	1.20	$3.55 \times 10^{-15}$	0.75
4	$-7.25 \times 10^{-15}$	1.16	$2.31 \times 10^{-4}$	0.71
5	$3.3 \times 10^{-3}$	1.25	$1.3 \times 10^{-3}$	0.70

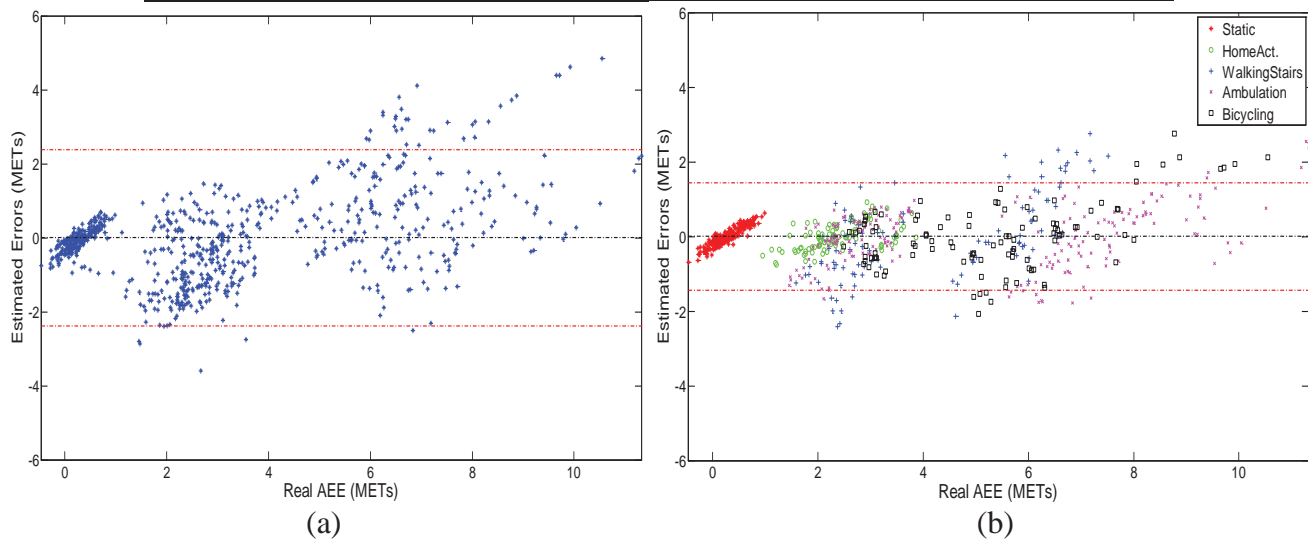


Fig. 3 Estimation errors of all activity categories by the regression models: (a) Without classification. (b) With classification. The middle line represents the estimated mean value. The dashed lines represent margin of twice the standard deviation.

Table 4 Average estimation error of AEE for each activity category with and without activity classification using the classifier combing the quartic regression model.

Activity categories	Errors of AEE (METs)		Without classification		With classification	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Static	-0.47	0.38	$6.7 \times 10^{-4}$	0.24		
Home activities	0.04	0.61	0.32	3.20		
Walking Stairs	-0.67	2.23	0.21	1.51		
Ambulation	-0.23	1.20	0.14	1.00		
Bicycling	2.05	1.24	0.16	0.97		

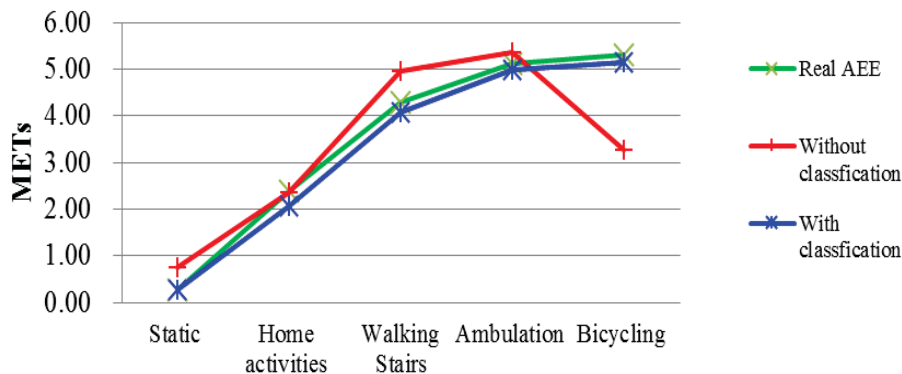


Fig. 3 Average MET estimation for each activity category with and without activity classification using the quartic regression model.

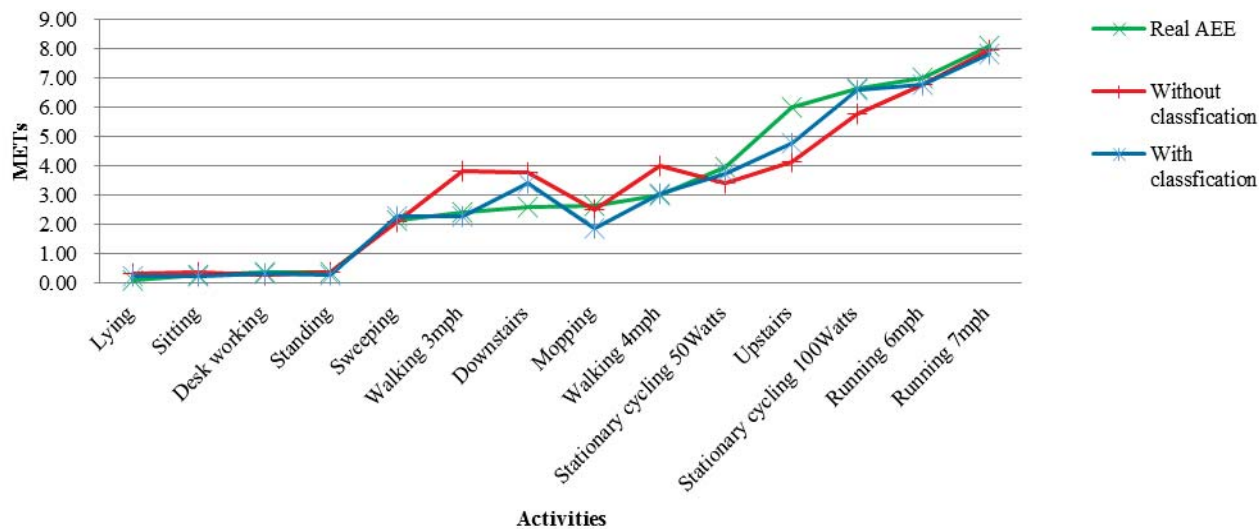


Fig. 4 Average MET estimation for each activity with and without activity classification using the quartic regression model.

However, combining the activity classifier to estimate METs could increase the estimation accuracy ranged from 0.31 METs to 1.23 METs for these six activities and basically maintained certain estimation accuracy for the rest. Thus, the above results verified that respectively conducting MET estimation according to different activity categories could improve the estimation accuracy of the system.

## 5 Conclusions

This paper presents of algorithms for physical activity classification and MET mapping regression model construction using a wearable physical activity sensor system. The proposed activity classification algorithm composes of the procedures of acceleration acquisition, signal preprocessing, and feature generation. The algorithm is capable of translating time-series acceleration signals into important feature vectors. The algorithm first extracts the time- and frequency-domain features from the acceleration signals, and the features are sent to a trained probabilistic neural network (PNN) for classification. Finally, the count and the user's characters such as height and weight are the features to construct MET mapping regression models to estimate activity energy expenditure. In our experiments, the average classification accuracy of five categories of physical activities for physical activity classification is 95.33%. The average error of MET estimation without and with activity classification is  $-7.25 \times 10^{-15} \pm 1.16$  METs and  $2.31 \times 10^{-4} \pm 0.71$  METs, respectively. In addition, the average errors of estimation with classification for walking stairs and bicycling are  $0.21 \pm 1.51$  METs and  $0.16 \pm 0.97$  METs, respectively. They are better than the single regression model whose error of walking stairs and bicycling are  $-0.67 \pm 2.23$  METs and  $2.05 \pm 1.24$  METs, respectively. The experimental results can validate that the performance of energy expenditure estimation with activity classification is improved more significant than without activity classification.

## 6 References

- [1] I. M. Lee, H. D. Sesso, and R. S. Paffenbarger Jr, "Physical activity and coronary heart disease risk in men does the duration of exercise episodes predict risk," *Circulation*, vol. 102, no. 9, pp. 981-986, 2000.
- [2] B. A. Calton, R. Z. Stolzenberg-Solomon, S. C. Moore, A. Schatzkin, C. Schairer, D. Albanes, and M. F. Leitzmann, "A prospective study of physical activity and the risk of pancreatic cancer among women (United States)," *BioMed Central Cancer*, vol. 8, no. 1, pp. 1-9, 2008.
- [3] S. Schulz, K. R. Westerterp, and K. Brück, "Comparison of energy expenditure by the double labeled water technique with energy intake, heart rate, and activity recording in man," *American Society for Clinical Nutrition*, vol. 49, no. 6, pp. 1146-1154, 1989.
- [4] G. Plasqui, A. Joosen, A. D. Kester, A. H. C. Goris, and K. R. Westerterp, "Measuring free-living energy expenditure and physical activity with triaxial accelerometer," *Obesity Research*, vol. 13, no. 8, pp. 1363-1369, 2005.
- [5] E. M. Tapia, "Using Machine Learning for Real-time Activity Recognition and Estimation of Energy Expenditure," Ph.D. dissertation, Massachusetts Institute of Technology, 2008.
- [6] S. J. Strath, D. R. Bassett, A. M. Swartz, and D. L. Thompson, "Simultaneous heart rate-motion sensor technique to estimate energy expenditure," *Medicine & Science in Sports & Exercise*, vol. 33, no. 12, pp. 2118-2123, 2001.
- [7] A. G. Brooks, S. M. Gunn, R. T. Withers, C. J. Gore, and J. L. Plummer, "Predicting walking METs and energy expenditure from speed or accelerometry," *Med. & Sci. in Sports & Exercise*, vol. 37, no. 7, pp. 1216-1223, 2005.
- [8] K. Y. Chen and D. R. Bassett, "The technology of accelerometry-based activity monitors: current and future," *Medicine & Science in Sports & Exercise*, vol. 37, no. 11, pp. S490-S500, 2005.
- [9] S. E. Crouter, J. R. Churilla, and D. R. Bassett Jr, "Estimating energy expenditure using accelerometers," *Eur. J. of Applied Physiology*, vol. 98, no. 6, pp. 601-612, 2006.
- [10] D. R. Bassett Jr, B. E. Ainsworth, A. M. Swartz, S. J. Strath, W. L. O'Brien, and G. A. King, "Validity of four motion sensors in measuring moderate intensity physical activity," *Medicine & Science in Sports & Exercise*, vol. 32, no. 9, pp. S471-480, 2000.
- [11] M. P. Rothney, M. Neumann, and A. Béziat, "An artificial neural network model of energy expenditure using nonintegrated acceleration signals," *Eur. Journal of Applied Physiology*, vol. 103, no. 4, pp. 1419-1427, 2007.
- [12] R. S. Rawson and T. M. Walsh, "Estimation of resistance exercise energy expenditure using accelerometry," *Med. & Sci. in Sports & Exer.*, vol. 42, no. 3, pp. 622-628, 2010.
- [13] C. A. Dorminy, L. Choi, S. A. Akohoue, K. Y. Chen, and M. S. Buchowski, "Validity of a multisensor armband in estimating 24h energy expenditure in children," *Med. & Sci. in Sports & Exer.*, vol. 40, no. 4, pp. 699-706, 2008.
- [14] G. Rodriguez, L. Michaud, L. A. Moreno, D. Turck, and F. Gottrand, "Comparison of the TriTrac-R3D accelerometer and a self-report activity diary with heart-rate monitoring for the assessment of energy expenditure in children," *British J. of Nutrition*, vol. 87, no. 6, pp. 623-631, 2002.
- [15] G. Welk, S. Blair, K. Wood, S. Jones, and R. Thompson, "A comparative evaluation of three accelerometry-based physical activity monitors," *Med. & Sci. in Sports & Exer.*, vol. 32, no. 9, pp. S489-S497, 2000.
- [16] A. M. Swartz, S. J. Strath, D. R. Bassett, W. L. O'Brien, G. A. King, and B. E. Ainsworth, "Estimation of energy expenditure using CSA accelerometers at hip and wrist sites," *Medicine & Science in Sports & Exercise*, vol. 32, no. 9, pp. S450-S456, 2000.
- [17] P. S. Freedson, E. Melanson, and J. Sirard, "Calibration of the computer science and applications, Inc. accelerometer," *Medicine & Science in Sports & Exercise*, vol. 30, no. 5, pp. 777-781, 1998.
- [18] S. E. Crouter, and D. R. Bassett Jr, "A new 2-regression model for the Actical accelerometer," *British Journal of Sports Medicine*, vol. 42, no. 3, pp. 217-224, 2008.
- [19] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, pp. 109-118, 1990.





## **SESSION**

# **NOVEL COMPUTER ALGORITHMS AND SYSTEMS + DRUG DISCOVERY + DATA MINING + NEW FRAMEWORKS FOR PREDICTION AND EARLY DETECTION OF DISEASES**

**Chair(s)**

**TBA**



# ECG analysis tool for heart failure management and cardiovascular risk assessment

J. Henriques<sup>‡</sup>, T. Rocha<sup>†</sup>, S. Paredes<sup>†</sup>, R. Cabiddu<sup>\*</sup>, D. Mendes<sup>‡</sup>, R. Couceiro<sup>‡</sup>, P. Carvalho<sup>‡</sup>

<sup>‡</sup>CISUC, Universidade de Coimbra, Coimbra - {jh, diana.sxm, rcouceir, carvalho}@dei.uc.pt

<sup>†</sup>Instituto Politécnico de Coimbra, Departamento de Engenharia Informática e de Sistemas, Coimbra, Portugal - {teresa, sparedes}@isec.pt;

<sup>\*</sup>DEIB, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy - ramona.cabiddu@mail.polimi.it

**Abstract-** This work presents the algorithms developed for the electrocardiogram (ECG) analysis, able to characterize the major parameters and events with clinical relevance for heart failure management and cardiovascular risk assessment. In particular, it includes algorithms for ECG pre-processing, ECG delineation, atrial fibrillation and ventricular arrhythmias detection, ST segment deviation and heart rate variability analysis.

This Matlab tool was originally developed in the HeartCycle FP7 European project and updated during the current CardioRisk project. The final goal of FP7 HeartCycle project was the management of heart failure patients based on a regular home monitoring of vital signals and other patient's parameters using wearable sensors. By means of this strategy it is possible to continuously assess patient's status, arrhythmic events and symptoms progression, thus enabling physicians to take the appropriate interventions, whenever abnormal situations are detected or foreseen. The CardioRisk project addresses the coronary artery disease, and the development of personalized algorithms for cardiovascular risk assessment, in particular using information obtained from heart rate variability analysis.

**Keywords** – Computational intelligence, ECG analysis, Heart failure, arrhythmias.

## I. INTRODUCTION

According to recent studies [1], non-communicable diseases accounted for almost 50% of the global disease burden. Among these, the highest incidences are for chronic degenerative diseases, namely cardiovascular diseases, responsible for approximately 50% of the disease burden [2]. Between the cardiovascular diseases, the coronary artery disease (CAD) and the heart failure (HF) are a major public health concern and the cause of considerable morbidity and mortality [3]. In fact, cardiovascular diseases are the world's primary cause of death, responsible for 17.1 million deaths per year. In particular, HF is a growing epidemic for which, despite clinical advances, mortality rates continues to be high. The progression of HF originates recurrent hospitalizations (acute decompensation events) and, even though the symptoms are reduced, the patient's cardiac function continually deteriorates [4].

The use of tele-monitoring technologies, together with adequate diagnosis and prediction methodologies, plays a decisive role in the conception of new eHealth solutions. In practice, it is possible the continuous monitoring of the

patient, the prompt diagnosis and the early prediction of critical events. Moreover, the information collected through the continuous tele-monitoring of patients during long periods creates new opportunities in the diagnosis of a disease, its evolution, and prediction of possible complications. In parallel, the risk assessment, i.e. the evaluation of the probability of occurrence of an event given the patient's past and current exposure to risk factors, is critical to improve diagnosis and prognosis [3]. As a result, it is clinically recognized that the research and development of practical and accurate CV risk assessment tools are of fundamental importance.

In this context, the analysis and the understanding of the ECG, in particular the extraction of relevant information and presence of abnormal rhythms, is central in the diagnosis of heart failure and in the development of cardiovascular risk assessment tools. In effect, the ECG plays a key role in monitoring and diagnosing of the cardiac diseases in general, being one of most valuable signals to characterize the cardiovascular status of the patient. The main goal of this work is to present a set of ECG algorithms for the characterization of parameters and events clinical recognized as fundamental for the HF and CAD management. The package was originally developed and integrated in the decision support system module of the HeartCycle project, and was recently improved in the context of CardioRisk project addressing personalized models for cardiovascular risk assessment.

The structure of this work is as follows: in section 2, the phases involved in the algorithms are presented as well as the most relevant aspects related with the algorithms. Section 3 presents the validation results using public datasets and in section 4, some conclusions are drawn.

## II. ALGORITHMS

### A. ECG analysis: main phases

Similarly to a common data mining procedure [5], the analysis and diagnosis of physiological signals comprises five main stages, as depicted in Figure 1: *i*) acquisition; *ii*) pre-processing; *iii*) data transformation or feature extraction; *iv*) modeling and validation; *v*) evaluation and interpretation.

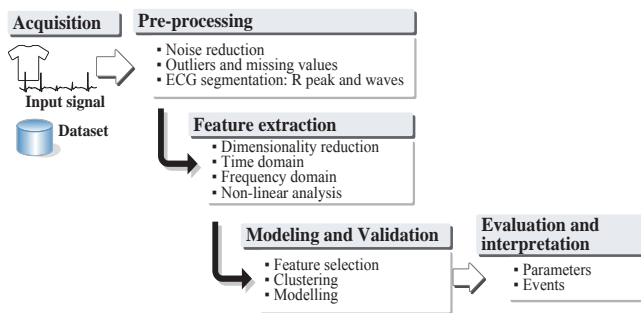


Fig 1. Steps involved in the process of ECG analysis and diagnosis.

The first phase involves the collection of physiological signals by means of adequate sensors. The second phase (pre-processing), in general includes operations applied to the raw data, in order to prepare it for further analysis. These operations typically include noise reduction, as well as the handling of outliers and missing values. The third step is, probably, the most important stage of the diagnosis process, and aims to transform the pre-processed data to a reduced set of meaningful parameters, able to characterize the data and holding a high discriminative value for the modeling process. This stage is often called feature extraction or dimensionality reduction phase, and may comprise dimensionality reduction techniques, as well as time-domain, frequency-domain, time-frequency domain and non-linear dynamic methods to compute the required features and parameters. The main goal of fourth phase is to build a model using the resulting of transformed data (features). Moreover, in case these features are redundant or its number is high, an additional feature reduction step can be implemented, aiming to obtain a compact subset between all the features. Among the possible alternatives, the following modelling approaches can be highlighted [6]: *i*) clustering models, to group objects using the concept of similarity, *ii*) classification models, to categorize objects into predefined classes, *iii*) regression models, to describe the relationship of a dependent variable from a set of independent variables, commonly used for prediction purposes. Finally, the last phase, includes operations for the quantitative assessment and analysis of the modeling process. Among the several performance measures possible to be used, the sensitivity and the specificity are examples of two statistical measures well established in the clinical field.

An ECG analysis package was developed and implemented during the HeartCycle project, which was integrated in the decision support system of the HF management platform. The architecture is modular and enables the detection of the most significant cardiac arrhythmias in HF management, including atrial fibrillation (AF) and ventricular arrhythmias (VA), as well as the detection of deviation in the ST segment. Recently, the HRV module of this package was improved in the context of the

cardioRisk project, which addresses the coronary artery disease and, in particular, the development of personalized risk assessment tools. In effect, existing risk assessment tools do not include HRV. Thus, further investigations are needed to establish which HRV parameters should be incorporated in existing tools to improve risk stratification that consists in one of the main goals of the cardioRisk project.

## B. Algorithms

### 1) Pre-processing

Among the numerous that have been proposed in literature, two particular methods have shown remarkable results, and are especially adapted to deal with noise artifacts and capable to operate in an online approach (or near real time approach), which are of major importance in tele-monitoring contexts. These two approaches are based on digital wavelet transforms [7] and multi-scale morphological operators [8]. In effect, the cyclical behavior of the ECG signal, together with its spectral components, reflected in the different frequency bands and scales can be distinguishable using multi-resolution decomposition by means of wavelet transforms or by applying morphological transform at different scales.

As result, in the ECG analysis performed: *i*) the baseline wander and noise removal was implemented using a wavelet approach [9]; *ii*) for the segmentation the method proposed by Sun, [8] with some adaptations was implemented. In effect, using morphological analysis, the most important fiducial points have been determined, enabling to characterize QRS complex, P and T waves, as well as the relevant intervals based on those waves identification. In particular the following parameters were computed:

*Segmentation*: P wave: P onset, P peak and P offset indexes; QRS complex: Q onset, Q peak, R peak, S peak and S offset indexes; T wave: T onset, T peak and T offset indexes.

*Intervals*: RR, heart rate (bpm), PR interval (s), corrected QT interval (s), Q wave width (s), Q peak height, R peak height, QRS complex duration (s) and corrected JT interval.

### 2) AF: Atrial fibrillation

From the clinical perspective the key characteristics of an AF episode is the absence of P waves before the QRS-T complex, which presents a *sawtooth like* pattern along the cardiac cycle, and the irregularity of the RR intervals. Thus, the proposed strategy makes use of the three principal physiological characteristics of AF, applied by cardiologists in their daily reasoning: *i*) P wave absence/presence, *ii*) heart rate irregularity and *iii*) atrial activity analysis. This knowledge-based approach has the advantage of increasing interpretability of the results to the medical community, while improving detection robustness [10].

A total of 6 features,  $f_i$ ,  $i = 1, \dots, 6$ , have been computed to address the above three characteristics. The features  $f_1, f_2$  and  $f_4$  are time-domain features; the feature  $f_6$  is a frequency-domain feature; the  $f_3$  and  $f_5$  features are computed using non-linear measures.

*P wave detection:* ( $f_1$ ) · the P wave absence/presence was quantified by measuring the linear correlation of each P wave to a normalized P wave model, created using common P waves extracted from the Physionet QT Database [11].

*Heart rate variability:* ( $f_2, f_3, f_4$ ) · in a first phase the RR interval sequence was modelled as a three-state Markov process, being each interval classified as one of the three states (short, regular or long), and characterized by its state transition probability matrix [12]. Three features were computed based on the transition probability, on the entropy of the distribution and on the Kullback-Leibler divergence measure between current window and model distribution.

*Atrial activity analysis:* ( $f_5, f_6$ ) · in a first step the method proposed by [13], was followed to cancel the QRS-T components from the ECG signal. Then the spectrum content of the obtained atrial activity was assessed by computing Kullback–Leibler divergence between each window spectrum and the current spectrum and a normalized AF spectrum model derived from the Physionet AF dataset [14].

### 3) PVC: Premature Ventricular Contractions

Most of the algorithms reported in literature for PVC detection are based on features derived from the QRS complex, independently from the surrounding ECG morphological characteristics. However, patients can exhibit several physical, clinical and cardiac conditions, which can affect the ECG morphology in numerous ways. For example, a wide QRS complex may be considered a non-PVC (normal beat) in one patient, while it may suggest the presence of a PVC in another patient [15]. The proposed algorithm approaches this problem by assuming that measurements extracted from PVC shape characteristics can be compared to normal, patient specific ECG beat characteristics and that these exhibit inter-individual resilience. Thus, in order to capture patient specific ECG characteristics, for each beat the measurements are compared with those extracted from the neighboring beats. A total of 13 features,  $f_i, i=1, \dots, 13$ , have been computed, mainly to characterize the QRS complex shape and its spectral components.

*Time domain features:* ( $f_1, \dots, f_8$ ) · the first group of features assesses the shape of the QRS complex, P and T waves. The first four relate the QRS complex (duration, area, center of mass and amplitude); the fifth computes the correlation between the actual P wave and a P wave template in order to decide about the presence/absence of the P wave; the T wave shape is characterized by its peak localization and deflection (sixth and seventh features). The RR regularity is used in the computation of the eighth feature.

*Morphological operators:* ( $f_9, f_{10}$ ) · two features are based on the ECG signal's morphological derivative, to estimate slopes before and after each R peak.

*Spectral information:* ( $f_{11}, f_{12}, f_{13}$ ) · the last three features are based on the frequency spectrum content, namely the entropy to assess the concentration of the current spectrum

and the Kullback–Leibler divergence between each normalized spectrum and the average of all spectrums.

### 4) VT: Ventricular tachycardias

The approach assumes that the fundamental differences in the physiologic origins of normal rhythm and VT, can be discriminated via time ECG shape together with power spectral density analysis. Moreover, the most significant features were selected from an original set of features (found in literature as well as developed during this work) [16], [17], [18] total of 11 features,  $f_i, i=1, \dots, 11$ , have been computed to discriminate between normal signals and VT episodes.

*Time domain features:* ( $f_1, \dots, f_6$ ) · the first feature estimates the morphology of the signal, computing the amount of time that each beat peaks is above or below a given threshold [19]. The second feature basically assess the heart rhythm [20]. The next four features assessing of small and high derivatives in the ECG signal, enabling to detect abnormal signal amplitudes and slopes [17].

*Spectral information:* ( $f_7, f_8, f_9$ ) · the energy contained in different frequencies (three ranges) was used as an approach for characterizing the ECG signal. The PSD was evaluated by windowing segments of signal, computed using the Welch's method.

*Non-linear features:* ( $f_{10}, f_{11}$ ) · one feature employs a non-linear transform, in particular the multiplication of backward differences [21], providing an estimation of extreme variations in the ECG. The other estimates the spatial filling index, computed from the ECG phase space reconstruction diagram [12].

### 5) ST deviation

The ST segment deviation is a measure computed as the difference between the isoelectric point (after the P wave) and the amplitude of the J point (segment of the ECG that presents a stable behavior, between the end of the QRS complex and beginning of the T wave). This amplitude, designated as ST segment deviation, is decisive in the assessment of the ischemic condition.

The algorithms implemented to evaluate ST segment deviation follow basically two stages.

1. *Baseline removal:* First, ECG signal is broken into cardiac cycles and a baseline removal process is applied to each individual interval. The main goal of this step is to guarantee that the isoelectric line is coincident with zero line, to facilitate ST segment shift evaluation.

2. *J Point evaluation:* The second stage involves several measures of the aimed deviation. In effect, the literature shows a great variety of approaches to assess this ECG feature. Four measurements of ST deviation are available. This way, the person analyzing the ST segment deviation, has three different values to support the decision making. The first three were chosen from literature, whose details are presented below. A new algorithm was developed and implemented based on Wigner Ville transform.

### 1. Baseline removal

Based on R peaks localization, the entire ECG signal is broken into cardiac cycles using the average of the distances between consecutive R peaks. Each cardiac cycle is then submitted to a process of baseline removal using Wolf's method [22], as shown in Figure 2.

This method starts to determine the initial and final heights (H1 and H2) of the interval, using the average of first five samples and the average of last five samples, respectively. Then, the line segment connecting H1 to H2 is subtracted from the ECG, originating a corrected signal in terms of baseline.

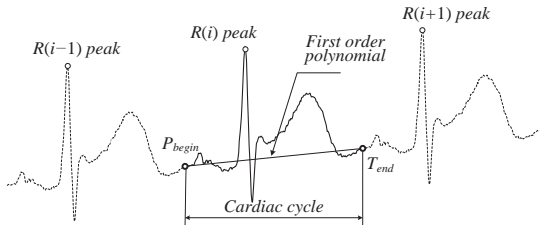


Fig 2. Base line removal (beat a beat approach)

### 2. J point estimation

Several alternatives can be found in literature. The first algorithm, proposed by [23], measures ST amplitude in the point localized 104 ms after the R peak. The second algorithm, introduced by [24], considers ST deviation 80 ms after the J point or, in case of sinus tachycardia (heart rate > 120 bpm), 60 ms after the referred point. This approach has the disadvantage of depending on J point accurate detection. The third method, used by [25], measures ST segment deviation in a point that depends on heart rate.

The method proposed here is based on a time-frequency analysis, showing notable results in what concerns the robustness of the J point identification. It is recognized that time-frequency methods are especially adequate for the detection of small transient characteristic hidden in the ECG, such as the ST segment. Thus, other approach for the estimation of ST deviation was based on a time-frequency analysis, in particular using the Wigner-Ville (WV) transform. The WV distribution is a time-frequency representation that considers a time analytical signal. Regarding the ECG, the equivalent analytic signal of the initial real signal  $x(n)$  was obtained by adding to the real signal its Hilbert transform  $H[.]$  as the imaginary part, equation (1)

$$y(n) = x(n) + jH[x(n)] \quad (1)$$

The WV transform fundamentally describes a signal, simultaneously in time ( $t$ ) and in frequency ( $f$ ), by considering an autocorrelation function of a signal, as (2).

$$W_X(nT, f) = 2T \sum_{k=-L}^L X(n+k) X^*(n-k) w(k) w^*(-k) e^{-j4\pi f k} \quad (2)$$

In the last previous equation,  $T$  represents the sampling period and  $w$  is a sliding window, symmetrical and with finite-length duration. In an analogy to the STFT, in the WVD the window is basically a shifted version of the signal itself.

The WVD is estimated by comparing the signal information with its own information at other times and frequencies.

The basic idea followed here consists in the division of the time frequency map into characteristic areas and, within each specific area, perform the evaluation of particular characteristics. With respect to ST estimation two time bands and one frequency band was considered. Regarding time band, the areas considered were those on the left (isoelectric line) and on the right (ST segment) of the R peak (assumed to be previously determined). For each time band it is expect to determine regions where there is no signal activity (isoelectric line, interval between the end of P wave and the begin of QRS complex, and ST segment, interval between the end of QRS complex and the begin of T wave). Thus, for those time bands, high frequency band were considered and, in particular, the region where high frequency components presents minimum values. Figure 3 depicts this idea, were an electrocardiogram and its corresponding high time-frequency components are shown (between 0.5 and 1.0, half of the normalized range). By evaluating the minimum of the sum of the high frequency components in each time band, isoelectric and J points can be obtained. Having determined these points ST deviation is straightforward estimated, as the difference between J and isoelectric values.

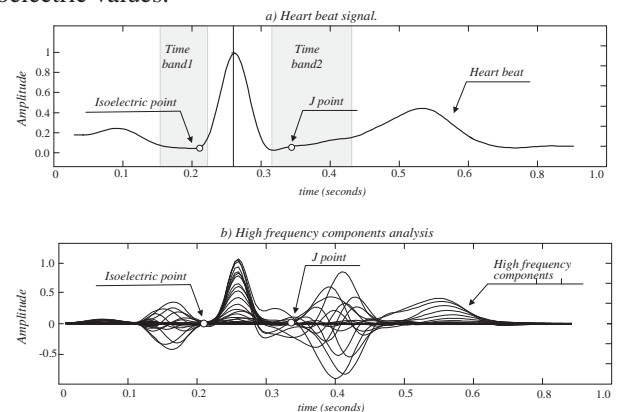


Fig. 5 ST segment deviation. a) electrocardiogram, isoelectric and J points b) frequency components (Wigner-Ville transform).

### 6) HRV: Heart Rate Variability

Various measures of HRV have been proposed in literature, which can generally be subdivided into time domain, frequency domain and non-linear measures. The algorithms implemented here to determine these measurements follow common approaches found in literature [26], [27] and no special effort was made to derive new measurements. The HRV analysis was essentially developed in the context of cardioRisk project and had a fundamental contribution of POLIMI [26]. The following parameters are available:

**Time Domain:** mean: mean of RR intervals; SDNN standard deviation of RR intervals; SDDSD standard deviation of the differences between heart beats (DHB); RMSSD root mean square of the DHB; NN50 number of RR intervals that fall within 50 milliseconds; pNN50 percentage of total NN50.

*Frequency domain:* PSD frequency content (Burg and Welch method are available); pVL percentage of very low frequency content [0 - 0.04]; pLF percentage of low frequency content [0.04 - 0.15]; pHF percentage of high frequency content [0.15 - 0.40]; rLF ratio pLF/pHF.

*Non-linear:* For each patient, Sample Entropy (SE), binary Lempel-Ziv Complexity (LZC1), ternary Lempel-Ziv Complexity (LZC2), short term Detrended Fluctuation Analysis (DFA1), long term Detrended Fluctuation Analysis (DFA2), Fractal Slope (FS), and the two Pointcaré Plots indices (PP1 and PP2) were implemented.

### III. CLASSIFICATION AND VALIDATION

#### A. Classification

The accuracy of a classifier is, obviously, highly dependent on the number of classes to be categorized. Clearly, with only two classes each classifier is able to provide a superior classification result, due to the lower complexity of the problem. This fact has justified the design of a distinct classifier (neural network) for each specific task (AF, PVC and VT). The proposed classifiers consists of feedforward neural networks with sigmoidal type activation functions, being its parameters, i.e., the weights and the bias trained using the Levenberg-Marquardt algorithm [28].

#### B. Data sets

To assess the performance of the classifiers a subset of the Physionet MIT-BIH databases was used, namely AF: atrial fibrillation database (AF) [29]; PVC: arrhythmia database [30]; VT: malignant arrhythmia database (VA) [30] and Creighton University ventricular tachyarrhythmia database (CV) [31]. For the assessment of the ST deviation the European ST dataset was used [24].

#### C. Validation

##### 1) ECG delineation

The ECG segmentation algorithm validation has been performed using all 105 records from MIT-QT Database. Record lead configurations most similar to MLII have been chosen for testing the algorithm. Table I shows the SE-sensibility and PP-positive predictivity results regarding ECG segmentation and intervals computation.

##### 2) PVC

The PVCs' detection algorithm validation has been performed using 46 of 48 Physionet QT database records [11]. Non MLII lead configurations records have been removed from the training and testing datasets, preserving coherence in the morphological characteristics of ECG records. 1965 PVCs and 11250 normal QRS complexes from the aforementioned dataset, compose the training dataset. Validation was performed using all dataset records.

##### 3) AF

To validate the proposed AF detection algorithm, 23 records from Physionet AF were used (lead MLII).

Respectively 19161 and 29893 windows of 12 seconds, corresponding to AF and non AF episodes, compose the training dataset. Validation has been performed using all 23 dataset records (238321 and 59785 AF and non AF episodes, respectively).

#### 4) Ventricular Arrhythmias

A data base of 51 signals was created, involving the three ECG signal classes (normal sinus rhythm, VT and VF). For MVA and CVT data sets, the number of windows was 420 (35 minutes) and 102 (8.5 minutes), respectively.

#### 5) Segmentation and arrhythmias results

The results obtained for the ECG segmentation and the arrhythmias detection are presented in Table II.

TABLE I  
CLASSIFICATION PERFORMANCE

	AF	PVC	VT (VA)	VT (CV)
Sensitivity	92.2	96.3	90.7	91.8
Specificity	91.4	99.1	95.0	96.9

#### 6) ST deviation

A truly validation process could not be done. In fact, the available databases in this area, namely, the European ST-T Database and the Long-Term ST Database, were created to be used for evaluation of algorithms that detect or differentiate between ischemic ST episodes, axis-related non-ischemic ST episodes, etc. This is not the case of the present algorithm, which only considers discrete values of the ST segment deviation without further processing. For this reason, a correlation analysis was carried out. The average results obtained are presented in the Table II.

TABLE II  
CORRELATION FOR THE ST SEGMENT DEVIATION MEASUREMENT

Method	Correlation	Records
Taddei's method	0.512	e0105, e0213, e0403, e0107, e0305, e0405, e0111, e0409, e0113, e0411, e0115, e0119, e0413, e0121, e0415, e0127, e0501, e0123, e0129, e0515, e0125, e0417, e0139, e0601, e0147, e0603, e0151, e0607, e0605, e0159, e0609, e0163, e0161, e0203, e0817, e0613, e0205, e0615, e0207, e0801, e0303, e0211, e0103, e0305,
Pang's method	0.575	
Akselrod's method	0.546	
Wigner Ville	0.576	

#### D. Discussion

The performance of the algorithms that can be actually compared are presented in Table I, revealing the significant discrimination performances. In effect for the arrhythmias detection the sensitivity and specificity values are perfectly in accordance with the state of the arte that can be founded in

literature. With respect to the ST segment the correlation values were computed showing the adequate results of the proposed strategy based on the time-frequency Wigner-Ville transform. The heart rate variability is based on well established parameters, thus without need to be comparable.

#### IV. CONCLUSION

In this work the integrated ECG analysis algorithm platform developed for the HF management product concept of HeartCycle and for the cardiovascular risk assessment in cardioRisk project was introduced. The proposed architecture is modular and enables the simultaneous detection of the most significant parameters and cardiac arrhythmias in HF and CAD management, such as AF, PVC, VT and VF detection, as well as heart rate variability analysis.

The validation of the algorithms was based on public databases. Classification results show that the proposed approach can be used to discriminate between different types of arrhythmias with state of the art accuracy. However, to effectively evaluate the developed algorithms, their performance has to be tested in real conditions. Under these circumstances, some modules must probably be improved to maintain/increase the obtained sensitivity and specificity and to assure robustness to changes in real measurements conditions (including noise and misunderstand events).

#### Acknowledgements

This work was supported by cardioRisk (PTDC/ EEI-SII/2002/2012) and FP7-iCIS (CENTRO-07-ST24-FEDER-002003).

#### V. REFERENCES

- [1] WHF (2009); World Heart Federation; Time to Act: The Global Emergency of Non-Communicable Diseases; Report on 'Health and Development: Held Back by Non-Communicable Diseases'.
- [2] Nugent (2008); R. Nugent; Chronic Diseases in Developing Countries, Health and Economic Burdens; Ann. N.Y. Acad. Sci. 1136: 70–79.
- [3] McMurray and Pfeffer (2005); J. McMurray, M. Pfeffer; Heart failure; Lancet, 365, 1877-1889.
- [4] Gheorghiadu et al (2005); M. Gheorghiadu, L. DeLuca, G. Fonarow, G. Filippatos, M. Metra, G. Francis; Pathophysiologic targets in the early phase of acute heart failure syndromes; Am J Cardiol 96, 11G -17G.
- [5] Sow et al (2013); D. Sow, D. Turaga, M. Schmidt; Mining of sensor data in Healthcare: a survey; Aggarwal (ed.), Managing and Mining Sensor Data, Springer Science Business Media New York, Chapter 14, 459-503.
- [6] Rokach (2010); L. Rokach; A survey of Clustering Algorithms; Data Mining and Knowledge Discovery Handbuuk, 269-298.
- [7] Martinez et al (2004); J. Martinez, R. Almeida, S. Olmos, A. Rocha, P. Laguna; A wavelet-based ECG delineator: evaluation on standard databases; IEEE Trans. Biomed. Eng., 51, 4, 570–581.
- [8] Sun et al (2005); Y. Sun, K. Chan, S. Krishnan; Characteristic wave detection in ECG signal using morphological transform; BMC Cardiovascular Disorders 5:28 (2005).
- [9] Couceiro (2008a); ECG analysis and atrial fibrillation detection; MSc Thesis, University of Coimbra.
- [10] Carvalho et al (2012); P. Carvalho, J. Henriques, R. Couceiro, M. Harris, M. Antunes, J. Habetha; Model-Based Atrial Fibrillation Detection; Chapter 5 - ECG Signal Processing, Classification and Interpretation; Ed: Adam Gacek, Witold Pedrycz; 99-133, ISBN:978-0-85729-867-6.
- [11] PhysionetQT; <http://www.physionet.org/physiobank/database/qtdb/>; MIT-BIH QT database
- [12] Tratnig (2005); R. Tratnig; Reliability of new Fibrillation Detection algorithms for automated External Defibrillators; PhD Dissertation, Technische Universitaet Graz.
- [13] Sanchez et al (2002); C. Sanchez, J. Millet, J. Rieta, F. Castells, J. Ródenas, V. Ruiz; Packet Wavelet Decomposition: An Approach for Atrial Activity Extraction; IEEE Computers Cardiology, 29:33-36
- [14] PhysionetAR (2014); Physionet; [Physionet:http://www.physionet.org/physiobank/database/mitdb/MIT-BIH\\_Arrhythmia\\_Database](http://www.physionet.org/physiobank/database/mitdb/MIT-BIH_Arrhythmia_Database).
- [15] Couceiro et al (2008c); R. Couceiro, P. Carvalho, J. Henriques, M. Antunes; On the detection of premature ventricular contractions; EMBC-08, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, Canada, August 20-24.
- [16] Henriques et al (2007); J. Henriques, P. Carvalho, P. Gil, A. Marques, T. Rocha, B. Ribeiro, M. Antunes, R. schmidt, J. Habetha; Ventricular Arrhythmias Assessment; EMBC -2007, 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Lyon, France, August 23-26, 2007.
- [17] Henriques et al (2008); J. Henriques, P. Carvalho, M. Harris, M. Antunes, R. Couceiro, M. Brito, R. Schmidt; Assessment of Arrhythmias for Heart Failure Management; phealth2008 - International Workshop on Wearable Micro and Nanosystems for Personalised Health; Valencia, Spain, May 21-23, 2008.
- [18] Rocha et al (2008); T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Antunes; Phase space reconstruction approach for ventricular arrhythmias characterization; EMBC-08, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, Canada, pp 5470-5473.
- [19] Tian and Tompkins (1997); L. Tian, J. Tompkins; Time domain based algorithm for detection of ventricular fibrillation; Proceedings of the 19 Int. Conference IEEE/EMBS Oct 30-Nov 2, Chicago, USA.
- [20] Jekova et al (2004); I. Jekova, G. Bortolan, I. Christov; Pattern Recognition and Optimal Parameter Selection in Premature Ventricular Contraction Classification; IEEE Comp.in Cardiology; 31: 357-360.
- [21] Kunzmann et al (2002); U. Kunzmann, G. Schochlin, A. Bolz; Parameter extraction of ECG signals in real-time; Biomed Tech(Berl). 4, 2:875-8.
- [22] Wolf, A. Automatic Analysis of Electrocardiogram Signals using Neural networks, (in Portuguese), PUC-Rio, Ms. Thesis, nº 0210429/CA2004.
- [23] Akselrod, S., Norymberg, M., Peled, I., Karabelnik E., Green, M. S. (1987). "Computerised Analysis of ST Segment Changes in Ambulatory Electrocardiograms", Medical and Biological Engineering and Computing, v. 25, p. 513-519.
- [24] Taddei A, Distanto G, Emdin M, Pisani P, Moody G B, Zeelenberg C and Marchesi C, The European ST Database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography Eur. Heart J. 13 1164–72, 1992.
- [25] Pang L, Tchoudovski I, Bolz A, Braecklein M, Egorouchkina K and Kellermann W 2005 Real time heart ischemia detection in the smart home care system 27th Annu. Int. Conf. Eng. Med. Biol. Soc., 2005. IEEE-EMBS 2005
- [26] Cabiddu et al (2013); R. Cabiddu, S. Mariani, J. Henriques, S. Cerutti, A. Bianchi; Non-linear Indices of Heart Rate Variability in Heart Failure Patients during Sleep; XIII Mediterranean Conf. MBEC, IFMBE Proceedings 41, 690-693, Seville, Spain, September, 25-28.
- [27] Tarvainen , M, Niskanen, J: Kubios HRV Analysis, version 2.0 beta, University of Kuopio, Kuopio, Finland, 2006. (<http://bsamig.uku.fi/>).
- [28] Hagan and Menhaj (1994); M. Hagan, M. Menhaj; Training Feedforward Networks with the Marquardt Algorithm; IEEE Trans. NN, 5, 6, 989-993.
- [29] PhysionetAF; <http://www.physionet.org/physiobank/database/afdb/>; MIT-BIH Atrial Fibrillation Database.
- [30] PhysionetAR; [http://www.physionet.org/physiobank/database/mitdb/MIT-BIH\\_Arrhythmia\\_Database](http://www.physionet.org/physiobank/database/mitdb/MIT-BIH_Arrhythmia_Database).
- [31] PhysionetCVT; <http://www.physionet.org/physiobank/database/cudb/>; MIT-BIH Creighton University ventricular tachyarrhythmia database.
- [32] PhysionetVF; <http://www.physionet.org/physiobank/database/vfdb/>; MIT-BIH malignant arrhythmia database.



# Understanding Medical Named Entity Extraction in Clinical Notes

Aman Kumar<sup>1</sup>, Hassan Alam<sup>1</sup>, Rahul Kumar<sup>1</sup>, Shweta Sheel<sup>1</sup>  
<sup>1</sup>BCL Technologies, San Jose, CA

**Abstract** - *Clinical notes contain extensive knowledge about patient medical procedures, medications, symptoms etc. In this paper we present an integrated approach to processing textual information contained in the clinical notes. We extract three major medical entities namely symptoms, medication and generic medical entities from patient discharge summaries and doctors notes from the I2B2 dataset. Quick access to structured information of these entities may help medical professionals in providing better and cost-effective care.*

**Keywords:** *Medical Named Entity Extraction, Machine Learning, Natural Language Processing*

## 1 Introduction

In recent years, various medical facilities and individual care providers have embraced electronic medical record (EMR) or some other version of electronic data management system. Understanding various facets of patient history is critical to speedy and economical treatment.

Clinical notes have been analyzed in greater detail to harness important information for clinical research and other healthcare operations, as they depict rich, detailed medical information. In this paper we describe a machine learning system which extracts medical named entities of three categories namely symptom, medication and generic medical condition. In this study we do not analyze text in the bio-medical journals or research papers. We analyze clinical text from doctor's notes and records that are generated during patient interviews. Clinical notes pose challenge for natural language processing in that they contain short phrases, abbreviations, acronyms etc. Sometimes there are instances of ungrammatical constructions in these notes.

This paper is organized as follows. First we provide a brief overview of various medical named entity extraction techniques and tools. We then describe the machine learning method that we use to build the medical named entity extraction system. Then we go over the experiment design followed by a section on results and discussion.

### 1.1 Related Studies

Named entity extraction is a type of information retrieval which focuses on identifying instances i.e., names of various types of entities. For example, cancer would be an instance of disease; swelling would be an instance of symptoms and so on. One of the earliest NER models was based on decision tree [1]. In this paper [1], Sekine used features such as part-of-speech tags extracted by a morphological analyzer, character based information and specialized dictionary. This system was developed for Japanese. Another early work was done by Bikel, Schwartz and Weischedel [2]. Authors used Hidden Markov Model (HMM) to identify named entity. Primary features like bi-gram and orthographic features like word case, word shape etc. were used. Borthwick [3] in his PhD thesis used maximum entropy (MaxEnt) algorithm.

McCallum and Li [4] developed Conditional Random Fields based algorithm to extract NER in coNLL-2003 shared task competition. Sarawagi and Cohen [5] propose a semi Markov CRF (Conditional Random Field) algorithm for named entity extraction. Cohen and Sarawagi [5] further extended the semi Markov model with use of dictionary and notion of similarity function. Naidu and Sekine [6] provide wide overall survey of NER research.

Aranson [7] developed MetaMap to map bio-medical concepts from Unified Medical Language System (UMLS). FriedMan et al. [8] developed a NLP system based on MetaMap which extracts various entities from clinical notes such as temporal information, corresponding codes by matching with UMLS etc. They extract concepts in semantic form based on various predefined frames. Minard et al. [9] presented and compared multiple approaches based on domain-knowledge and machine-learning techniques to Medical Entity Recognition. They show that the hybrid approach based on both machine learning and domain knowledge obtains the best performance. Li, Schuler and Savova [10] have used both CRF and SVM based for model extraction of *disorder* in clinical text. In this paper Authors have extracted a dictionary from SNOMED-CT [11]. SNOMED-CT is a map of concepts and relationships. In total it contains almost 360,000 and 1 million relationships. It is classified into categories such as procedure, body part etc. Several such models has been developed which uses variation of statistical models aided with dictionary based system. Wang & Patrick [12] present a cascaded system to do NER on clinical notes. Their system consists of two of CRF and SVM respectively. Patrick & Li [13] developed a CRF based

methods to extract medication from clinical text. Meystree et al. [14] provide an overview of recent developments in clinical information retrieval field.

## 2 Methodology

Our medical named entity extraction system is modeled with a CRF model. In the literature it has been consistently demonstrated that CRF models are best performing models for sequential labeling. For the present study we extended and modified the Stanford NER package [15]. We prefer Stanford NER package because it can be seamlessly integrated with existing NLP tool suite such as parser and morphological analyzer.

CRFs are undirected graphical models which can be interpreted as conditionally trained finite state machines. While based on the same exponential form as maximum entropy models, they have efficient procedures for complete, non-greedy finite-state inference and training.

Here are the definitions:

Let  $G=(V,E)$  be a graph such that  $Y = (Y_v)_{v \in V}$ , so that  $Y$  is indexed by the vertices of  $G$ . Then  $(X,Y)$  is a conditional random field in case, when conditioned on  $X$ , the random variables  $Y_v$  obey the Markov property with respect to the graph:  $p(Y_v|X, Y_w, w \neq v) = p(Y_v|X, Y_w, w \sim v)$ , where  $w \sim v$  means that  $w$  and  $v$  are neighbors in  $G$ .

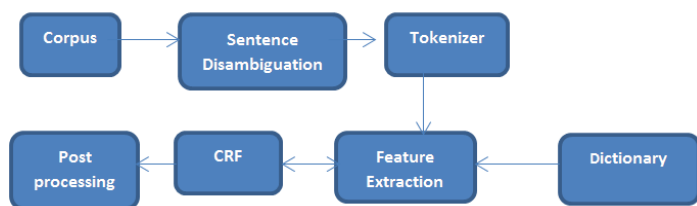
If the graph  $G = (V, E)$  of  $Y$  is a tree, the conditional distribution over the label sequence  $Y = y$ , given  $X = x$ , by fundamental theorem of random fields is:

$$p_{\theta}(y|x) \propto \exp \left( \sum_{e \in E, k} \lambda_k f_k(\theta, y|_e, x) + \sum_{v \in V, k} \mu_k g_k(v, y|_v, x) \right)$$

Where

$\theta = \lambda_1, \lambda_2, \dots, \lambda_n, \mu_1, \mu_2, \dots, \mu_n$  are weights to be estimated by the model.

The architecture of our system is given below.



The I2B2 corpus was annotated with the following tags.

1. MNE-S: refers to medical named entity symptom.
2. MNE-M: refers to medical named entity medication.

3. MNE: This tag refers to generic medical named entity which may not qualify under previously defined categories but are useful to physicians in understanding patient clinical record.

The I2B2 corpus was annotated in Begin-Inside-Outside (BIO) format. We noticed that several clinical notes had a lot of discrepancies with respect to sentence boundary. So we used Stanford CRF model to first break the corpus into sentences and then these sentence were broken into tokens. We curated an extensive dictionary of medical conditions, symptoms and medications using SNOMED-CT.

### 2.1 Feature Extraction

A brief overview of features extracted is described below.

1. Word based features: We extracted N-gram features to understand word context. We used a window of  $N=1, 2, 3, 4$ . We also used word shape as a feature. This was appropriate to understand some tokens like temperature and concepts like dosage, frequency etc. We also used morphological features such as prefix and suffix of words.
2. Semantic Knowledge: Known acronyms and synonyms were extracted.
3. Orthographic features: Spelling variations and spelling corrections were extracted.
4. Parse tree features: We used Stanford parser to extract semantic features. We tagged tokens with part-of-speech tags. We extended these features in N-gram fashion ( $N=1,2,..$ ). We also introduced distance of tokens from numeric quantifiers, if present in a sentence.
5. Dictionary based features: We used Boolean features - if present token is a medication or a symptom or a specific medical condition etc.
6. To further reduce noise related to abbreviations such as q.i.d, Dr. etc., we developed a lexicon. Such tokens were represented with specific features during model development.
7. Extensive regex based features to identify time entities were used. Presence of such entities indicates context for specific medical events (such as pain in the night etc.) in the clinical notes.
8. Character level features: Presence of special characters like /, @, -, :, mixed cases, alphanumeric characters in a word etc were harnessed. Each such occurrence was represented with a unique Boolean flag.
9. Aspell dictionary check: We introduced a special Boolean flag if the word was present in Aspell English dictionary.

### 3 Experiment and Evaluation

We extracted corpus from I2B2 dataset (<https://www.i2b2.org/NLP/DataSets/>). In total we sampled 2100 sentences from the corpus. Our motive for sampling was that we wanted a system that was not overly representative of a specific entity such as symptom. We wanted a balanced system that had equal representation of all entities under investigation. These sentences were then annotated by a physician in BIO format for each entity. We split the corpus in 70/30 ratio for training and testing respectively. Training of the model was done with 10-fold cross validation.

Table 1 shows the evaluation scores of our system. We use standard Recall, Precision, and F-score to measure the performance of the system.

**Table 1.** Experiment Results

	MNE*	Wang & Patrick* [12]	MNE-M*	Patrick & Li* [13]	MNE-S
<b>Precision</b>	82.34%	83.30%	91.34%	89.62%	79.41%
<b>Recall</b>	71.89%	76.78%	70.84%	81.38%	62.65%
<b>F score</b>	76.76%	79.91%	79.79%	84.88%	70.04%

\*, # represent same category.

In the results presented above, we show that our system has done extremely well in simultaneous extraction of multiple entities from clinical notes. The comparable systems although perform marginally better, they only extract a single type of entity unlike ours where we extract three entities – MNE-M, MNE-S, and base MNE. Also, existing systems do not extract MNE-S (Medical Named Entity – Symptom).

### 4 Conclusions

In this paper we presented an integrated machine learning system to extract three major types of medical named entities. We believe such an integrated system is more practical than systems that extract entities in isolation.

Our future efforts will be toward building a larger granular set of medical entity extraction, which will further improve multi-class medical entity extraction system. Such a system can be used for preparing medical reports based on patient's records such as patient discharge summary and doctor's notes.

### 5 References

- [1] Sekine, S. 1998. Nyu: Description of the Japanese NE System Used For Met-2. In Proc. Message Understanding Conference
- [2] Bikel, D. M., Schwartz, R., & Weischedel, R. M. An algorithm that learns what's in a name. *Machine learning* 34, no. 1-3 (1999): 211-231.
- [3] Borthwick, A. A maximum entropy approach to named entity recognition. PhD diss., New York University, 1999.
- [4] McCallum, A & Wei L. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4, pp. 188-191.
- [5] Sarawagi, S. & Cohen, W. W. Semi-markov conditional random fields for information extraction. In *Advances in Neural Information Processing Systems*, pp. 1185-1192. 2004.
- [6] Cohen, W. W., & Sarawagi, S. Exploiting dictionaries in named entity extraction: combining semi-markov extraction processes and data integration methods. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 89-98. ACM, 2004.
- [6] Nadeau, D. & Sekine, S. A survey of named entity recognition and classification. *Linguisticae Investigationes* 30, no. 1 (2007): 3-26.
- [7] Aronson AR. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program
- [8] Friedman, C., Shagina, L., Lussier, Y., & Hripcsak, G. (2004). Automated Encoding of Clinical Documents Based on Natural Language Processing. *Journal of the American Medical Informatics Association : JAMIA*, 11(5), 392–402.
- [9] Minard AL, Ligozat AL, Ben Abacha A, et al. Hybrid methods for improving information access in clinical documents: concept, assertion, and relation identification. *J Am Med Inform Assoc.* 2011; 18(5):588–93
- [10] Li, D., Kipper-Schuler, K., & Savova, G. Conditional random fields and support vector machines for disorder named entity recognition in clinical texts. In Proceedings of the workshop on current trends in biomedical natural language processing, pp. 94-95. Association for Computational Linguistics, 2008.
- [11] Bos, L., and K. Donnelly. "SNOMED-CT: The advanced terminology and coding system for eHealth." *Stud Health Technol Inform* 121 (2006): 279-290.
- [12] Wang, Y. & Patrick, J. Cascading classifiers for named entity recognition in clinical notes. In Proceedings of the workshop on biomedical information extraction, pp. 42-49. Association for Computational Linguistics, 2009.
- [13] Patrick, J., & Li, M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. *Journal of the American Medical Informatics Association*, 17(5), (2010):524-527.
- [14] Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. Extracting information from textual documents

in the electronic health record: a review of recent research. *Yearb Med Inform* 35 (2008): 128-44.

[15] Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 363-370. Association for Computational Linguistics, 2005.

# Mining Frequent Embedded Sub trees

B.Chandra

Indian Institute of Technology, Delhi, India

**Abstract** - The area of mining structured data is of large interest in real life applications. It has found its applications in finding frequently occurring structures in Chemical compounds, Protein structures, Bio Chemical compounds etc. In this paper we have proposed an alternative approach to mining embedded sub trees in a collection of rooted ordered labeled sub trees. The proposed technique was compared with TreeMiner algorithm suggested by Zaki [12]. The frequent embedded sub trees mined using the proposed approach were same as those generated using TreeMiner, while the time taken while mining the sub trees using both the approaches was comparable.

**Keywords:**

## 1 Introduction

Data mining has consistently evolved over the years and found its applications in various domains. It evolved from finding patterns in simple 2 dimensional datasets (ie transactional datasets) using techniques like association rule mining [1], sequence mining[2] to mining frequently occurring substructures in high dimensional datasets ( like Protein structures, chemical compounds etc) using graph mining [7],[8], [9] ,[11] and tree mining[3],[4],[5],[10],[12]. The applications of tree mining are found in bioinformatics, biological data from the field of DNA and protein analysis, common topological patterns in RNA structures, mining XML documents, Web logs etc. It can also be used for finding frequent substructures in a collection of several similar chemical compounds. These frequent substructures might be important for the activity of the compound. If an unknown chemical compound contains same substructure then this unknown chemical compound could also be used for the same purpose as that of existing compounds. Another example of application of tree mining is to find secondary structures in RNA molecules. To understand the similarity of different RNA molecules it is necessary to compare their structures instead of their sequences alone. As two different sequences can produce similar secondary structures, comparisons between secondary structures are necessary to lead to a better understanding of the functions of different RNA molecules.

In [12] an algorithm, based on enumeration trees called TreeMiner is proposed, to discover all frequent embedded sub trees, i.e., those sub trees that preserve ancestor-descendant relationships in a collection of rooted ordered trees. FREQT algorithm [3] discovers frequent rooted ordered sub trees in a

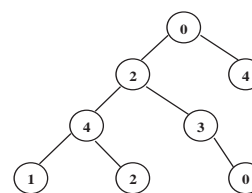
collection of trees. Algorithms for mining rooted unordered sub trees based on enumeration tree growing are presented in [3] & [5]. Canonical forms for rooted unordered trees were suggested in both the papers to overcome the problem of having multiple ordered trees corresponding to the same unordered tree. In [4] an Apriori-like algorithm, FreeTreeMiner is discussed, to mine all frequent free sub trees. Xiao et al presented the Path Join algorithm [10] which uses the FP-tree [6] data structure for mining maximal frequent sub trees.

The problem of finding frequent embedded sub trees has been discussed in depth by Zaki in [12]. In this paper we propose a novel alternative approach for mining frequent embedded sub trees in a collection of rooted ordered labeled trees. An alternative representation of the collection of trees has been evolved to facilitate a simpler approach for finding frequent sub trees. The frequent embedded sub trees generated using the proposed approach was compared with TreeMiner. The proposed approach was tested using large datasets and the performance was at par with the TreeMiner.

## 2 Proposed Algorithm

### 2.1 String Encoding of each tree in the database:

All the trees in the data base are converted into list format (refer figure 1). This list is called as horizontal list. The horizontal list is taken as an input for finding out frequent embedded sub trees.



$\mathcal{T}$  (T's String Encoding): 0 2 4 1 -1 2 -1 -1 3 0 -1 -1 4 -1

Fig. 1. An Example of a Tree in Database

The label number of the nodes of the tree is listed in a preorder manner with the exception that a minus 1 is inserted while back tracking each time during traversing the tree. Add the vertex label to  $\mathcal{T}$  (The string encoding of tree T) while traversing the tree T in a preorder manner and add a unique symbol  $-1 \notin L$  (where  $L = \{l_1, l_2, l_3, \dots\}$  is the set of labels) whenever we backtrack from a child to its parent. So here in the given example (refer Fig. 1) we start from the root add 0,2,4,1 to the string and then add  $-1$  as we backtrack and then

add 2 and so on. Figure 2 shows a collection of three trees in the Dataset DT

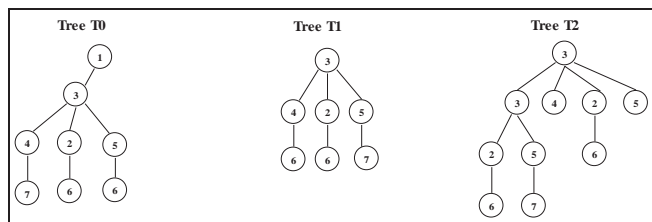


Fig. 2. Dataset DT

The string encoding for the above trees is given by:

T0 : 1 3 4 7 -1 -1 2 6 -1 -1 5 6 -1 -1 -1  
 T1 : 3 4 6 -1 -1 2 6 -1 -1 5 7 -1 -1 -1  
 T2 : 3 3 2 6 -1 -1 5 7 -1 -1 -1 4 -1 2 6 -1 -1 5 -1

2.2 Equivalence Classes:

If the prefix up to the  $(k-1)$ th node is common to two  $k$ -sub trees, then we say that the two sub trees belong to the same prefix equivalence class. Thus the members of the equivalence class differ only in the position of the last node.

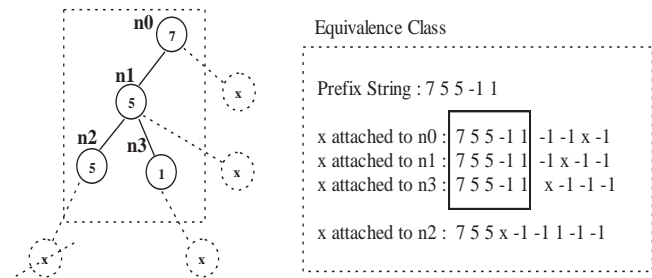


Fig 3. Prefix Equivalence Class

The Figure 3 shows the class template for a sub tree of size 5 with the  $[P]_4$ , prefix of size 4. It is observed that if node with label  $x$  is attached to nodes  $n0, n1$  and  $n3$  the prefix remains the same with string encoding  $7 5 5 -1 1$ . However if the node with label  $x$  is attached to node  $n2$  then  $[P]_4$  for the resultant tree of size 5 would be  $7 5 5 x -1$  which is different from that when  $x$  is attached to  $n0, n1$  and  $n3$ . Hence this tree does not belong to the same prefix equivalence class as the other three sub trees.

2.3 Computing  $F_1$

For each distinct item  $i \in T$ , the string encoding of  $T$ , we increment  $i$ 's count in a 1D array. All labels in  $F_1$  belong to the class with empty prefix, given as  $[P]_0 = [\emptyset] = \{(i, -1), i \in F_1\}$  and “-1” indicates that  $i$  is not

attached to any node. The total time for this step is  $O(n)$  per tree, where  $n = |T|$ . The item that has support greater than or equal to the minimum support is used for further candidate generation to obtain  $F_2$ . In the database DT (refer figure 2) the  $F_1$  elements are  $(2, -1), (3, -1), (4, -1), (5, -1), (6, -1)$  and  $(7, -1)$  with prefix  $[P]_0 = \{ \}$  for minimum support of 100%.

2.4 Computing  $F_k (k > 3)$

Horizontal String encoding for each tree is stored in a 2D vector  $D$  as shown in the example database DT. The vector  $D$  stores the Tree id along with the entire Tree also. The String encoding is scanned for each tree. The numbers are stacked up one on top of the other until a “-1” is encountered. On finding a “-1” in the string encoding we move to the next column of the row where the label was currently put. Now on we move down one step at a time in this column if continuous -1's occurs in the Horizontal list. On finding a value other than -1 we now start staking up the numbers one on top of the other till the next -1 occurs in the list. Repeat the procedure till the end of the Horizontal list.

```

Trans(Dataset DT)
{Initialize i,j,k,m to zeros
flag=1;
for each tree  $T_i$  in DT
while(!end of list)
if( $T_i[m] = -1$ )  $D[j][k] = T_i[m]$ ;
Increment j,k by one;
flag=1; else
if(flag=1) increment k by one;
flag=0; else decrement j by one;
end
end
end
increment D;
end}
    
```

Fig. ...

	Tree T0	Tree T1	Tree T2
J			
3	7 6 6		6 7
2	4 2 5	6 6 7	2 5 6
1	3	4 2 5	3 4 2 5
0	1	3	3
k	0 1 2	0 1 2	0 1 2 3 4

Fig. 4. Vector Representation of each Tree in Dataset DT

Now for example we consider the Horizontal list for Tree T0 of Dataset DT. The list contains 1, 3, 4, 7 and then a -1. Hence the value of j is incremented keeping the value of k constant and the values are filled into the matrix. When the -1 is interfaced the value of j = 3. On finding -1 the value of k is incremented by one. This value is incremented only once on the first occurrence of -1 after a non minus one value. Now here in the list we find the occurrence of an additional -1. Here we do not increment k but the value of j is decremented. The value of j is decremented each time a -1 is found consecutively. At this point the value of k is 1 and that of j is 2. The list now contains 2 and 6 before a -1. Hence now 2 is inserted at (j=2, k=1) and 6 is inserted at (j=3, k=1). This process is repeated till we reach to the end of the list. The conversion of horizontal list to a vector representation is carried out while finding the 1-Frequent items. The algorithm for finding frequent embedded sub tree of size greater than 2 is shown below:

```

Find_subtree(Vector D, Frequent k-1 elements, num_fk)
{
Mfk.clear();
for each tree t in D
  for each Frequent k-1 element f
if (size of tree>length of f)
  find the last index L(r,c) of f stored in Multimap Mf ;
  for each frequent one item f1 which has the same prefix f
if( the last number in the prefix is not equal to "-1")
if(f1 is in the same column c above the row r)
  insert the pair <f,f1> into Multimap Mfk
  for the first occurrence of f1;
store the position of the element "f f1" in Multimap Mf
end if
end if
if (the last number in the prefix is equal to "-1")
find the last index L(r,c) of "f minus
last two elements" stored in Multimap Mf
if ((f1 is in any column c' > c above the row r)&& D(r,
c')=null)
  insert the pair <f,f1> into Multimap Mfk for the first
occurrence of f1;
store the position of the element "f f1" in Multimap Mf
end if
end if
end
  else erase of the entries of tree t from Multimap Mf
  remove the tree t from Vector D
  end if
end
end
Frequent_subtrees=Frequent_subtrees U Frequent k-1
subtrees stored in Mfk
Mfk.clear()
Find_frequent(Mfk)// Find_frequent is a function for
finding frequent itemsets

```

```

Update_fk(fk-1,f1)
{ fk.push_back(strcat(fk-1,f1);}
Sub_Tree_miner(Dataset)
{ Find f1 as discussed in Section2a while generating the
Vector D
Update Multimap Mf by the key equal to root and
value equal to (0,0)
k=1;
while (fk.size!=0)
  Find_subtree(D , fk, k+1)
Display all the sub trees stored in Frequent_subtrees}

```

### 3 Results

We have used the synthetic data generation program by Zaki [12] which mimics website browsing behavior. The program first constructs a master website browsing tree, W, based on parameters supplied by the user. These parameters include the maximum fan-out F of a node, the maximum depth D of the tree, the total number of nodes M in the tree, and the number of node labels N. Multiple nodes in the master tree to can the same label. The master tree is generated using the following recursive process. Once the master tree has been created many subtrees of W are created as specified by the parameter T. To generate a sub tree a recursive process is followed starting at the root. A random number between 0 and 1 is generated to decide which child to follow, or to backtrack. If a branch has already been visited, then other unvisited branch is selected, or backtracking is carried out. The nomenclature used in naming the dataset is "T<size of dataset>F<max fan-out>D<max depth>.data". For example T10KF10D10.data means the dataset is for size 10000 with maximum fan-out of 10 and maximum depth of 10. For each dataset, the number of labels kept is 100 and the master tree was generated was of 10000 nodes. The results for minimum support values 0.01%, 0.02% and 0.03% are shown in Tables 1, 2 and 3 respectively.. It was observed that the number of frequent sub trees generated using both the algorithms were same.

**Table 1.** Results for datasets for minimum support = 0.01%

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
T1KF10D10	13	21	42	49	36	14	2	0			
T2KF10D10	14	26	60	80	70	35	9	1	0		
T3KF10D10	15	29	79	121	122	77	31	8	1	0	
T4KF10D10	18	42	117	203	239	191	105	39	9	1	0
T10KF10D10	20	58	198	421	621	649	479	242	78	14	1

**Table 2.** Results for datasets for minimum support 0.02%

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
T1KF10D10	9	11	16	12	5	1	0			
T2KF10D10	13	21	39	43	28	9	1	0		
T3KF10D10	13	22	46	58	48	24	7	1	0	
T4KF10D10	13	23	53	75	69	39	13	2	0	
T5KF10D10	13	23	56	85	87	60	28	8	1	0
T10KF10D10	16	36	103	185	227	186	98	30	4	0

**Table 3.** Results for datasets for minimum support 0.03%

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
T1KF10D10	7	8	11	8	2	0				
T2KF10D10	8	10	14	11	3	0				
T3KF10D10	12	20	38	45	31	12	2	0		
T4KF10D10	12	22	46	60	51	27	8	1	0	
T5KF10D10	12	22	48	67	62	37	13	2	0	
T10KF10D10	15	33	82	140	161	124	61	17	2	0

## 4 Conclusions

An alternate tree mining approach has been presented in this paper. It was observed on a number of synthetic datasets, that it produces the same sub trees as that generated by TreeMiner. The approach is a very simple and has an easy implementation using data structures like vector and multimap.

## 5 References

Number in square brackets (“[ ]”) should cite references to the literature in the main text. List the cited references in numerical order at the very end of your paper (under the heading ‘References’). Start each referenced paper on a new line (by its number in square brackets).

- [1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the Twentieth International Conference on Very Large Databases*, pages 487–499, Santiago, Chile, 1994.
- [2] R. Agrawal and R. Srikant. Mining sequential patterns. In *Proceedings of the 11th International Conference on Data Engineering*, Taipei, Taiwan, Mar. 1995. IEEE Computer Society Press. T. Asai, K. Abe, S. Kawasoe, H. Arimura, H. Satamoto, and S. Arikawa. Efficient substructure discovery from large semi-structured data. In *Proc. of the 2nd SIAM Int. Conf. on Data Mining (SDM'02)*, April 2002.
- [3] T. Asai, H. Arimura, T. Uno, and S. Nakano. Discovering frequent substructures in large unordered trees.

In *Proc. of the 6th International Conference on Discovery Science (DS'03)*, October 2003.

- [4] Y. Chi, Y. Yang, and R. R. Muntz. Indexing and mining free trees. In *Proc. of the 2003 IEEE Int. Conf. on Data Mining (ICDM'03)*, November 2003A. Full version available as Technical Report CSD-TR No. 030041 at <ftp://ftp.cs.ucla.edu/tech-report/2003-reports/030041.pdf>.
- [5] Y. Chi, Y. Yang, and R. R. Muntz. Mining frequent rooted trees and free trees using canonical forms. Technical Report CSD-TR No. 030043, <ftp://ftp.cs.ucla.edu/techreport/2003-reports/030043.pdf>, UCLA, 2003B.
- [6] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *2000 ACM SIGMOD Intl. Conference on Management of Data*, pages 1{12,2000.
- [7] A. Inokuchi, T. Washio, and H. Motoda. An apriori-based algorithm for mining frequent substructures from graph data. In *Proceedings of the 4th European Conference on Principles of Knowledge Discovery and Data Mining*, sep 2000.
- [8] M. Kuramochi and G. Karypis. Frequent subgraph discovery. In *Proceedings of the 1st IEEE Int'l Conference on Data Mining*, nov 2001.
- [9] D. Shasha, J. Wang, and R. Giugno. Algorithms and applications of tree and graph searching. In *Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 39–52, Madison, Wisconsin, June 2002.
- [10] Y. Xiao, J-F Yao, Z. Li, and M. Dunham. Efficient data mining for maximal frequent subtrees. In *Proc. of the 2003 IEEE Int. Conf. on Data Mining (ICDM'03)*, 2003.
- [11] X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002)*, 9-12 December 2002, Maebashi City, Japan, pages 721–724. IEEE Computer Society, 2002.
- [12] M. J. Zaki. Efficiently mining frequent trees in a forest. In *8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, July 2002.



## Detection of cardiac events in the context of a rehabilitation platform

T. Rocha, S. Paredes

*Instituto Superior de Engenharia de Coimbra, Portugal {teresa, sparedes}@isec.pt*

J. Henriques, P. de Carvalho

*Centro de Informática e Sistemas, Universidade de Coimbra, Portugal {jh, carvalho}@dei.uc.pt*

### Abstract

**Abstract** – This work presents the strategy for the early detection of cardiovascular events that was integrated in a rehabilitation platform developed during the FP7 European project HeartWays. The project addressed the development of methodologies able to predict the evolution trend of biosignals time series collected by telemonitoring systems, in order to support the early detection of critical events (such as hypertension and arrhythmic episodes).

The approach is based on the hypothesis that current and past measurements taken from a historical dataset can be the support for the estimation of biosignals future evolution. Two main phases are involved: a similarity analysis procedure to find a set of similar patterns in the historical dataset, and a prediction scheme that makes use of the obtained patterns to forecast the future evolution trend. The validation of the algorithms was performed using blood pressure and heart rate signals collected during the myHeart telemonitoring study.

**Keywords** – Early detection, trends prediction, similarity measures, wavelets transform.

### I. INTRODUCTION

Approximately 5% of all deaths in Europe are due to cardiovascular diseases, and more than 20% of all European citizens suffer from a chronic cardiovascular disease, such as arrhythmias, congestive heart failure and coronary artery disease [1]. Coronary artery disease (CAD) is caused by an accumulation of plaques within the walls of the arteries that supply the myocardium with oxygen and nutrients. After long periods of progression, some of these plaques may rupture and, along with the limited blood supply to the myocardium, may result in an heart attack (myocardial infarction), requiring urgent hospitalization.

Although it is clinically recognized that cardiac rehabilitation provides many benefits for patients, its effective implementation after hospitalization remains low [2]. The use of remote patient monitoring and treatment (RMT) systems, based on telemonitoring services that enable professionals to access and evaluate symptoms and status progression, offers a huge potential in the context of cardiac rehabilitation. In effect, through the continuous monitoring of vital signs (such as blood pressure, heart rate, and body weight) the status and the quality of life and care of the patient can be assessed, and the prediction of aggravations and exacerbations of its chronic condition evaluated.

The FP7 European project HeartWays, (FP7-SME-315659) [3], aims the development of advanced modular solutions for supporting cardiac patients in rehabilitation outside an hospital centre, with the aid of wearable sensors and intelligent algorithms that personalize the management and the follow up for patients and professionals. Figure 1 depicts the approach followed in the project, which consists of three main technological workpackages: WP1: Smart Monitoring Layer; WP2: Multiparametric Analysis Layer; WP3: Patient Support and Healthcare Management Layer.

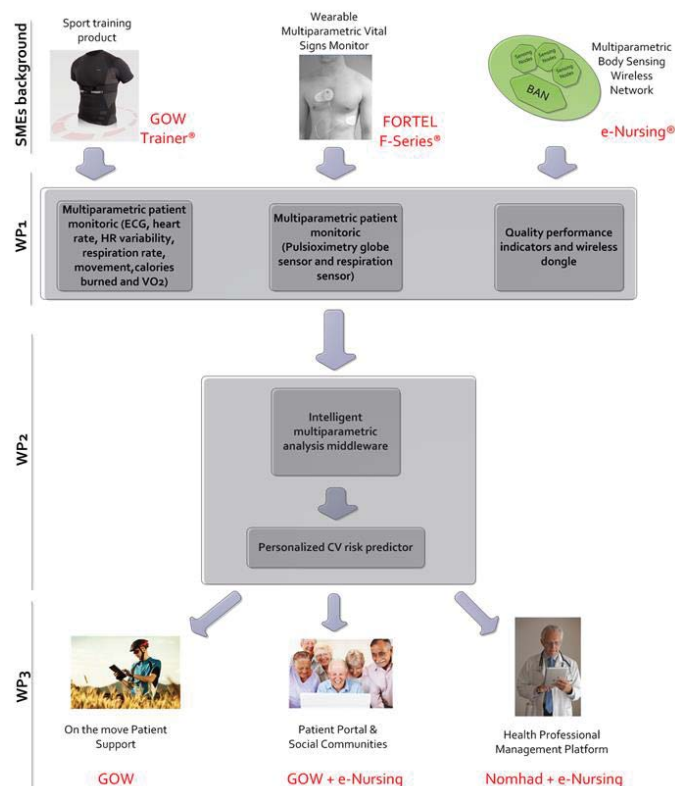


Figure 1: Schematic diagram of the HearWays project

Included in WP2, one of the modules comprises the development of methodologies for biosignals prediction, mainly to support the early detection of critical events. The specific biosignals to be addressed are daily collected by the system, namely blood pressure (BP) and heart rate (HR). The particular

events to be detected are hypertension and arrhythmic episodes, based on the evolution of BP and HR, respectively.

In terms of strategy, this module is founded on the hypotheses that the estimation of biosignals' future evolution can be supported on current and past measurements, captured from a historical dataset.

As result, from the research and practical perspectives, two major topics are addressed: *i*) how the trends in time series can be captured and compared; *ii*) how these trends can be used in the prediction process.

For the assessment of similarity between time series two main groups of algorithms can be identified: time domain and transform-based methods [4]. In the context of this work, the time-frequency analysis methods, included in the second group of algorithms, assume a particular interest. This is the case of the wavelet transform that produces features that describe properties of the time series both at various locations and at numerous time granularities, which is particularly important when dealing with the similarity assessment problem [5]. In particular, a specific Haar wavelet was used in the proposed similarity search scheme. Given its characteristics, the application of the wavelet transform was also considered for trend extraction and time series prediction. In effect, it provides a formal method to de-noise, de-trend, and decompose time series, capturing useful information at various resolution levels, so that the capacity of a forecasting model can be improved [6].

The structure of this work is as follows: section 2 presents the similarity analysis and prediction schemes. Section 3 discusses its application to heart rate and blood pressure signals for the detection of arrhythmic and hypertension episodes, using data collected during myHeart tele-monitoring study. Finally, in section 4, some conclusions are drawn.

## II. METHODOLOGY

This work is founded on the hypothesis that the estimation of biosignal's future evolution can be supported on current and past measurements taken from a historical dataset. In effect, the proposed methodology involves two main stages, as illustrated in Figure 2.

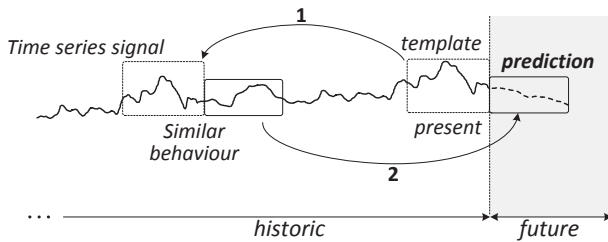


Figure 2: Prediction of biosignals evolution

*Similarity analysis procedure:* firstly, by means of a similarity analysis process, the selection of patients who display similar behaviors in their physiological time series is carried out;

*Estimation of future evolution trend:* then, the estimation of the biosignal's future values is performed, based on the similar time series identified in the first stage.

Basically, the process starts by considering the current signal to be predicted, designated here as the template,  $X(t) \in \mathbb{R}^{1,N}$ . Using the template and from a similarity analysis procedure, the set of the  $M$  most similar patterns  $\mathbf{X}(t) \equiv \{X_m(t) \in \mathbb{R}^{1,N}\}$ ,  $m=1, \dots, M$ , is identified. From these, the corresponding subsequent  $P$  future values,  $\mathbf{Y}(t) \equiv \{Y_m(t) \in \mathbb{R}^{1,P}\}$ , are straightforwardly obtained (known past values from historical dataset). Then, the known "future" evolution of the identified patterns,  $\mathbf{Y}(t) \equiv \{Y_m(t)\}$ , can be used in a prediction mechanism to estimate the future evolution of the current template,  $\hat{Y}(t) \in \mathbb{R}^{1,P}$ .

### A. Similarity Analysis Procedure

As previously referred, the first step of the prediction strategy consists in selecting patients that display similar behaviours in their physiological time series (historical dataset). To this end, algorithms able to find the segments of a time series that present the same dynamics of a given temporal template (indexing process) should be developed. The proposed methodology for evaluating the similarity between two physiological time series combines the Haar wavelet decomposition, in which signals are represented as linear combinations of a set of orthogonal basis, with the Karhunen-Loève transform, that allows for the optimal reduction of that set of basis. The similarity measure is based on the Euclidean distance, which is indirectly calculated by means of the linear combination coefficients of both time series. Furthermore, using an iterative algorithm for computing the referred coefficients, computational complexity of the method significantly decreases. Figure 3 illustrates the procedure.

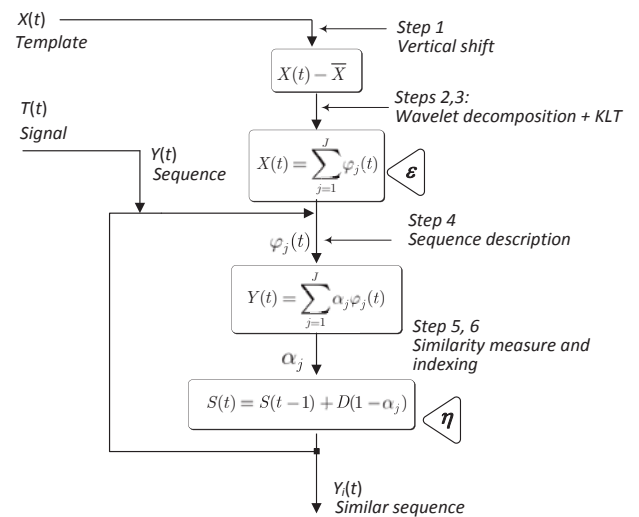


Figure 3: Similarity analysis scheme.

Follows a brief description of each step identified in the above scheme.

- *Step 1* - Vertical shift removal: to guarantee that similarity assessments are independent of variations in the vertical position, a vertical shift removal procedure is employed.
- *Step 2* - Wavelet decomposition of the template: to be compared with the time series is achieved by means of a set of orthogonal wavelet basis.
- *Step 3* - Optimal dimension reduction: based on the localization property of the wavelet basis, the ones that significantly reflect the dynamical patterns of the template are chosen to compose a reduced set of basis.
- *Step 4* - Sequence description: a subsequence of the signal to be compared with the template is described by means of the previous reduced set of basis. It is important to refer that this description does not involve a wavelet decomposition, but a simple computation of coefficients.
- *Step 5* - Similarity measure: the coefficients obtained from the template and subsequence description using the reduced set of basis, are employed to derive a similarity measure. This measure allows the interpretation as a trend evolution, as well as a percentage of the amplitude difference between the time series.
- *Step 6* - Subsequence indexing: based on the previous similarity measure, and using the particular Haar wavelet, an efficient iterative similarity indexing algorithm is proposed.

The parameters to be selected,  $\varepsilon \in \mathbb{R}^+$  and  $\eta \in \mathbb{R}^+$  correspond to:  $\varepsilon$ : controls the approximation error by determining the number of basis to be considered in the template decomposition;  $\eta$ : establishes if two signals that present the same behaviour are or not similar by thresholding the difference in amplitudes of the two series under comparison.

### B. Estimation of Future Evolution Trend

Figure 4 depicts the global scheme for the trend estimation of biosignals. It is composed of three main distinct phases: *i*) similarity analysis to identify patients who display similar behaviours in their physiological time series (previously presented); *ii*) multi-resolution decomposition of the time series retrieved from such patients,  $\mathbf{X}(t) \equiv \{X_m(t) \in \mathbb{R}^{1,N}\}$  and  $\mathbf{Y}(t) \equiv \{Y_m(t) \in \mathbb{R}^{1,P}\}$ ; *iii*) projection of the current patient data (template),  $X(t)$ , into the future,  $\hat{Y}(t)$ , by combining the optimal decomposition levels of the historic patterns  $Y(t)$ .

It is important to refer that the methodology presented here does not explicitly involve a model. In effect, it is based on the wavelet decomposition of the similar historical patterns, in order to derive an optimal future trend for the template. To achieve this goal two main steps are need: *i*) computation of the representative trends and; *ii*) computation of the optimal trends.

*Representative trends:* The first step involves the wavelet decomposition of the similar historical time series signals. Then, at each decomposition level, the obtained decompositions are combined to obtain a representative trend. The subtractive clustering method is used for this purpose.

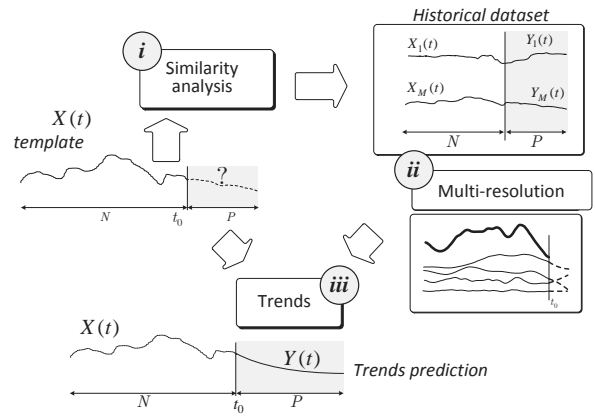


Figure 4– Proposed wavelet multi-resolution prediction methodology.

*Optimal trends:* The second step aims at the selection of a subset from the representative trends, designated as optimal trends, to be used in the prediction of the current time series. To achieve this goal, an optimization process involving the minimization of a set of distance-based measures is proposed. From this process, it is possible to quantify the aptness of each individual representative trend to integrate the optimal subset. Finally, the optimal trends are then straightforwardly extended to the future and aggregated to derive the global trend,  $\hat{Y}(t)$ .

## III. RESULTS

### A. Dataset

For the experiments, a private dataset resulting from the MyHeart European project was used [7]. The MyHeart project consisted in a home telemonitoring system that followed the health of heart failure patients, enabling intervention when appropriate. This was done by monitoring vital body signs with wearable technology, processing the measured data and giving recommendations (when appropriate) to the patient and professional users of the system. Using the measured data to give user feedback, the system “closed the loop” of measurements and therapy. This system was used in a clinical observational study carried out with 148 patients from six clinical centres in Germany and Spain. The trial had an enrolment phase of 9 months with 12 months of patient follow up. During the clinical study patients were requested to daily measure weight, blood pressure, and, using a vest, heart rate and bio impedance, as well as breathing rate and activity during the night by means of a bed sensor. Moreover, they were requested to complete each day two questionnaires of symptoms and mood/general well-being. From the 148 patients recruited, 102 (69%) were considered analysable, that is, with more than 30 days of telemonitoring measurements.

### B. Prediction Methodologies

Two groups of experiments were carried out. The first group assesses the capacity of the proposed wavelet multi-resolution scheme (here designated by WMM) in the trend prediction of

BP and HR signals. Moreover, the performance of this scheme is compared with other typical prediction strategies, namely a linear regression model, the autoregressive integral moving average model - ARIMA, and a non-linear regression model, the generalized regression neural network - GRNN. Other prediction method (AVP) simply considers the average value of predictive signals  $Y_m(t)$ , as an estimation for the prediction of  $Y(t)$ .

The second set of experiments selects patients with BP/HR values in a critical range (around the threshold of hypertension/tachycardia), and uses the previously estimated trend to determine the risk of hypertension/tachycardia. Specifically, the goal is to evaluate whether during the following week the BP/HR signal of a given patient evolves towards hypertension/tachycardia values or, on the contrary, is maintaining or decreasing to normal values.

### i. Parameters

With respect to the ARIMA model, the examination of the autocorrelation and partial autocorrelation functions of the differenced series, was used in the estimation of the order of the model  $ARIMA(n_a, d, n_c)$ . The parameters  $n_a, d$  and  $n_c$  identify, respectively, the number of autoregressive terms, the degree of differencing and the number of lagged forecast errors in the prediction equation. As result, the ARIMA structure was  $ARIMA(2,1,2)$ . The estimation of parameters was carried out with the  $armax(\cdot)$  Matlab command.

Regarding GRNN structure, it can be seen as normalized radial basis function networks, where there is a hidden unit centred at every training case. These units are called "kernels" and, usually, are probability density functions, such as Gaussian functions. The weights from the hidden to output layer are just the target values, so the output is simply a weighted average of the target values of the training cases, close to the given input case. As a consequence, the only parameters to be learned are the widths of the units. In the experiments using BP and HR signals, the width of the kernels was experimentally determined as  $\lambda=0.2$ . The  $newgrnn(\cdot)$  Matlab command was used to implement this neural model. Moreover, a different neural network had to be trained for each template.

With respect to AVP, the average prediction  $\bar{Y}(t)$ , of the identified patterns was computed using an weighted average, taking into account the similarity measure evaluated for each pattern.

The last approach (WMM) put into practice the proposed wavelet strategy, considering the following parameters:

*Similarity analysis:*  $N=32$ ,  $P=8$ , where  $N$  and  $P$  denote, respectively, the time intervals before and after the current time instant;  $M=5$ , number of patterns retrieved from the historical dataset;  $L=5$ , wavelet decomposition level.

*Selection of the optimal trends:* Number of decompositions considered in the optimal trend selection  $l=3,4,5,6$  (the details are the levels  $l=3,4,5$ ; the approximation is the level  $l=6$ ); the first two levels of detail ( $l=1,2$ ) were neglected;

conjunction and aggregation operators were, respectively, the  $maximum(\cdot)$  and the  $product(\cdot)$  operators.

### ii. Metrics

The accuracy of the forecasting methods was determined in terms of four performance metrics: *i*) the proposed similarity measure based on the wavelet decomposition+KLT (SWK), (1); *ii*) the Pearson's correlation coefficient (CORR), (2); *iii*) the normalised root mean squared error (NRMSE), (3) and *iv*) the mean absolute percentage error (MAPE).

$$SWK = S Y(t), Y(t) \quad t = N+1, \dots, N+P \quad (1)$$

$$CORR = \frac{\sum_{t=N+1}^{N+P} Y(t) - \bar{Y} \quad Y(t) - \bar{Y}}{\sqrt{\sum_{t=N+1}^{N+P} Y(t) - \bar{Y}^2} \sqrt{\sum_{t=N+1}^{N+P} Y(t) - \bar{Y}^2}} \quad (2)$$

$$NRMSE = \frac{1}{P} \frac{\sum_{t=N+1}^{N+P} Y(t) - Y(t)^2}{\sum_{t=N+1}^{N+P} Y(t) - \bar{Y}^2} \quad (3)$$

$$MAPE = \frac{1}{P} \sum_{t=N+1}^{N+P} \left| \frac{Y(t) - Y(t)}{Y(t)} \right| \quad (4)$$

In the previous equations,  $Y(t)$  is the actual BP/HR value,  $Y(t)$  is the forecasted BP/HR,  $\bar{Y}$  and  $Y$  are, respectively, the means of the actual and the estimated signals. The metrics NRMSE and MAPE were transformed into  $NRMSE = \exp(-\kappa_N NRMSE)$  and  $MAPE = \exp(-\kappa_M MAPE)$ , in order to guarantee that their values are in the range  $[0,1]$ . The parameters  $\kappa_N$  and  $\kappa_M$  are constants, respectively,  $\kappa_N = 0.25$  and  $\kappa_M = 10$ .

### iii. Statistical validation

Among the available parametric and nonparametric tests, the Friedman test is a nonparametric one that enables to perform multiple comparisons in experimental studies. This test (Friedman, 1937), (Friedman, 1940) is equivalent to ANOVA and is particularly adequate for machine learning studies when the assumptions (independency, normality and homoscedasticity) do not hold or are difficult to verify for a parametric test [8].

The objective of the Friedman test is to determine if it is possible to conclude, from a set of results, that there is a difference among the several methods. Basically, the Friedman test compares the average ranks  $R_j$  of each method, to decide about the null hypothesis, which states that "*Ho: all the algorithms behave similarly and thus their ranks  $R_j$  should be equal*". The Friedman statistics, is distributed according to  $\chi_F^2$ , with  $k-1$  degrees of freedom. From the computation of the corresponding *p-value*, the null hypothesis can be or not rejected at a given level of significance.

The Nemenyi test enables a pairwise comparison of the methods, based on the average ranks computed in the Friedman test. Basically, by means of the Nemenyi test, two methods can be significantly different at several levels, namely  $\alpha=1\%$ ,  $\alpha=5\%$ , or  $\alpha=10\%$ , if their average ranks differ at least the critical value. In this case ( $k=4$ ) the thresholds for the critical values are, respectively,  $CD_1=1.4675$ ,  $CD_5=1.2110$  and  $CD_{10}=1.080$ .

C. Trends Prediction

This section summarizes the results obtained from the application of the predictive WMM strategy to the prediction of blood pressure and heart rate signals.

1. Prediction of blood pressure

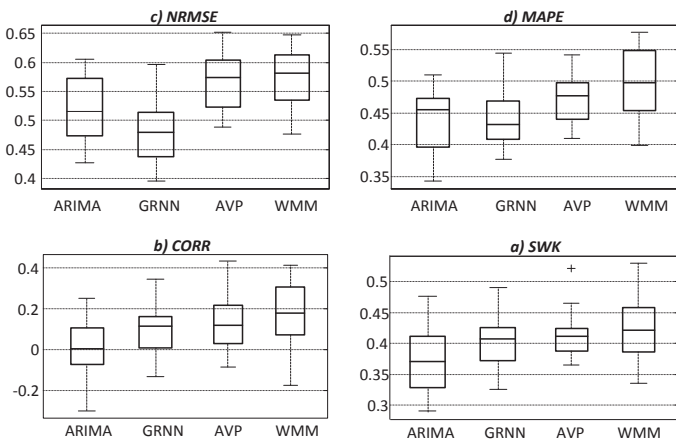


Figure 5 – BP: Comparison of the prediction methods.

Observing all the similarity measures in Figure 5 it is not possible to identify a method that clearly achieves superior results in comparison with the others. In global terms, it appears that the proposed method (WMM) is comparable or slightly superior to the others.

2. Prediction of heart rate

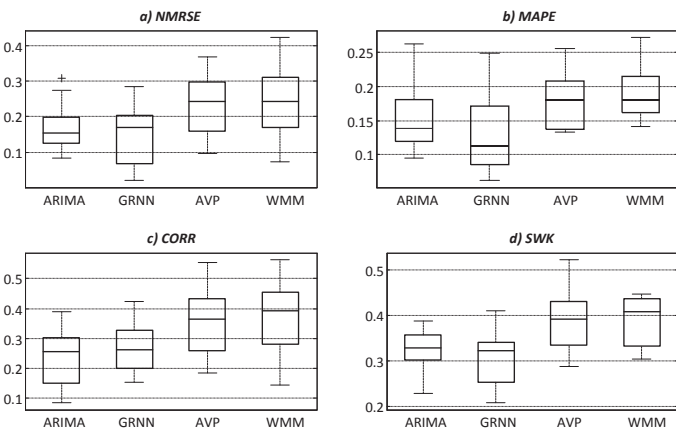


Figure 6 – HR: Comparison of the prediction methods.

Regarding the prediction of heart rate signals, Figure 6 shows the box-plots with the results corresponding the prediction, respectively.

As in the blood pressure case, by the observation of all similarity measures it is not possible to elect the best method. In global terms, it is possible to conclude that the performance of WMM is comparable to some of the methods and superior to others.

D. Detection of Events

1. Risk of developing hypertension episodes

A group of experiments was particularly applied to patients whose blood pressure values were in a critical range (around the threshold of hypertension). The main goal was to employ the trend prediction results to assess the hypertension risk of a patient. Specifically, the aim was to determine whether during the following week the blood pressure signal of that patient would evolve towards hypertension values or, on the contrary, would be maintained or decrease to normal values. Figure 7 illustrates this idea.

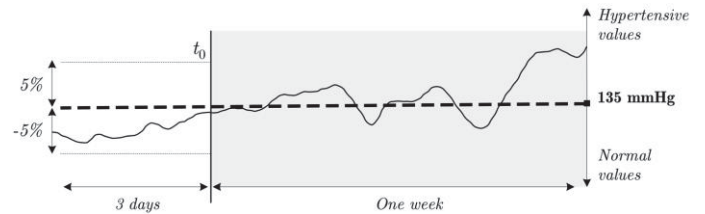


Figure 7 – Assessment of the hypertension risk.

The procedure started by identifying the patients that had recently shown blood pressure values in a critical range, more specifically, that had presented blood pressure values in the range  $[-5\%, +5\%]$  of the limit value of 135 mmHg during 3 consecutive days.

Then, for those patients, the blood pressure values of the following week were predicted. According to the percentage of values that were above the limit threshold (135 mmHg), the risk of the patient was assessed: if the percentage was higher than 75%, the patient was considered to be at risk of developing an hypertension episode condition; in the other case (less than 75%), the patient was considered to have no risk. Table I shows the discrimination capability of the method.

TABLE I. HYPERTENSION EVENTS

Predicted class	Actual class	
	No risk	In risk
No risk	68	3
In risk	6	18

Using MyHeart dataset, 150 experiments were performed using random templates, in which 35 BP signals exhibited values in the critical range. From these experiments, the obtained values of sensitivity (SE) and specificity (SP) were of, respectively, 85.7% and 91.8%. These values demonstrate the potential of the trend prediction strategy.

## 2. Risk of developing of arrhythmic episodes

A set of experiments was carried out particularly applied to patients whose heart rate values were in a critical range (around the threshold of tachycardia). The main goal was to determine whether during the following week the heart rate signal of a patient would evolve towards tachycardia values or, on the contrary, would be maintained or decrease to normal values.

The procedure is the same that was applied to hypertension case (HR limit value of 100 bpm during three consecutive days). The effectiveness of the proposed strategy was tested by selecting, from a set of 600 random templates, the ones that verified the referred requirement (to be in the critical range). In effect, 58 verified this condition: in 26 cases the patient presented risk of developing a tachycardia episode, and in 32 cases the patient revealed no risk. Table II shows the discrimination capability of the method.

TABLE II. ARRHYTHMIC EVENTS

	Actual class		
	No risk	In risk	
Predicted class	No risk	28	10
	In risk	4	16

To quantify the validity of the method, the sensitivity and specificity were determined, resulting in a SE of 61.5% and a SP of 87.5%.

Although it was not possible to compare these results with other works, considering that the prediction involved fully random templates, the obtained SE and SP values for both cases were very satisfactory. In effect, these metrics demonstrate the potential of the trend prediction strategy.

## 4. CONCLUSIONS

The main goal of Heartways project, (FP7-SME-315659), was the development of an advanced modular solution for supporting cardiac patients in rehabilitation outside an hospital center, with the aid of wearable sensors and intelligent algorithms that personalize the management and the follow up for patients and professionals. One of the objectives at the algorithm level (Multi-parametric Trends Analysis and Events Prediction Algorithms) was the development of methodologies for biosignal prediction, mainly to support the early detection of critical events (such as hypertension episodes based on the evolution of blood pressure, and arrhythmic episodes based on the evolution of heart rate).

This paper proposed a strategy based on wavelet decomposition for the prediction of biosignals, which goal was to estimate signals' future evolution trend. The capability of the proposed methodology was in a first phase compared with other common prediction mechanisms. Then, using the predicted values, the scheme was tested in the assessment of hypertension and arrhythmic events in patients whose blood pressure/heart rate values were in a critical range. For the effect, real data collected by the tele-monitoring study under myHeart project was used. The obtained values of sensitivity and specificity reveal the capacity of the strategy.

## REFERENCES

- [1] World Health Organization; *The Atlas of Heart Disease and Stroke*; Edited by J. Mackay and G. Mensah, 2004.
- [2] Reeves, G., D. Whellan; *Recent advances in cardiac rehabilitation*; *Current Opinion in Cardiology*, 25(6): 589-596, 2010.
- [3] Lazaro, et al; *Advanced solutions for Supporting Cardiac Patients in Rehabilitation*; <http://www.tsb.upv.es/proyectos/encurso/heartways/documentos/HeartWaysProjectBrochure.pdf>
- [4] Rocha, T., et al., *An effective wavelet strategy for the trend prediction of physiological time series with application to pHealth systems*, IEEE EMBC, Osaka, 2013.
- [5] Fryzlewicz et al; *Forecasting non-stationary time series by wavelet process modeling*; *ISM.*, v55, n4, 737-764, 2003.
- [6] Renaud et al; *Prediction based on a multiscale decomposition*; *Int. Journal of Wavelets, Multiresolution and Information Processing*, V. 1, N. 2, pp. 217-232, 2003.
- [7] Habetha, J; *MyHeart - A new approach for remote monitoring and management of cardiovascular diseases*; *IEEE EMBC 2006*.
- [8] Garcia et al. (2010); S. Garcia, A. Fernández, J. Luengo, F. Herrera; *Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power*; *Information Sciences: an International Journal archive*, Volume 180 Issue 10, May, 2010, pp 2044-2064.

# Improving Drug Discovery Process by Identifying Frequent Toxic Substructures in Chemical Compounds - A Graph Mining Approach

B.Chandra<sup>1</sup>, Shalini Bhaskar<sup>1</sup>

<sup>1</sup> Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016

**Abstract** - *Discovery of drug using computer modeling is one of the major challenges in contemporary medicine. Developing new therapeutic drugs is an expensive and time consuming process. Toxicity, caused by substructures that are carcinogenic in nature, is one of the important aspects that need to be explored during drug discovery. The use of in silico methods at an early stage of drug discovery can greatly reduce the time to test the drugs in laboratories. Identifying frequent substructures related to carcinogenic toxicity from a database of chemical compounds can be of great help in reducing the search space during drug discovery process. Graph mining algorithms can be used in order to identify substructures that occur frequently in a given compound database.*

**Keywords:** graph mining algorithms, drug discovery, substructures, carcinogenic

## 1 Introduction

New drug discovery is a challenging task since a single compound is declared safe for research on human beings among thousands of new compounds. There has been a phenomenal increase in the size of the chemical compound database in the recent years. This has changed the entire course of discovering and developing new drugs. Databases available for chemical compounds are being used as the initial point for screening candidate molecules and hence facilitate the pharmaceutical industry to manufacture more than one hundred thousand new compounds every year [23]. The use of efficient computer modeling techniques can be employed for drug discovery to reduce the search space. Compounds that are identified as promising are further investigated in the process of development wherein apart from the other parameters that include absorption, distribution, metabolism and excretion (ADME); their prospective toxicity is evaluated. This is a complex and costly process that often requires years before the compounds can be tested on human beings [9]. Further, about ninety percent of the initially chosen drug candidates fail to reach the market due to their toxic properties [26]. A compound can be declared toxic if it contains carcinogenic substructures, substructures that can

cause mutations, etc. This fact underlines the significance of determining probable toxic effects at an early stage in the process of development. Toxicity tests decide whether a candidate molecule is expected to create toxic effects in human beings. It generally involves using animal models at a pre-clinical phase. There is a greater need for competent *in silico* techniques to predict the toxicity of chemical compounds with the increase in number of biological targets and the demand for drug screening campaigns. Graph mining algorithms have been efficiently used for finding frequent substructures from database of chemical compounds that can cause toxicity.

Identification of frequent substructures that cause toxic effect in chemical compounds can greatly help the new drug discovery process since such substructures could be avoided in the new drugs. Graph mining facilitates the process of drug discovery by uncovering the chemical and biological characteristics of the drug. Thus, pharmaceutical industry can hope to reduce both the time and cost involved in the drug discovery process.

Graphs can be used for representing chemical compounds with atoms represented by the vertices in the graph and bond between the atoms as edges in the graph. Identification of frequent substructures is one of the predominant tasks performed by graph mining. Various graph mining algorithms have been discussed in the literature. AGM [16], a vertex growth method, follows apriori principle [1] for candidate generation procedure. FFSM [14] uses adjacency matrix for storing each isomorphic graphs. Code for each isomorph is generated by reading adjacency matrix. *Subdue* [15], an approximate mining algorithm, identifies frequent substructures and at the same time compresses graph dataset. *gSpan* [28] uses DFS tree for representing graph. A graph can have more than one DFS tree with DFS code associated with each DFS tree. Graph mining algorithms have been used in the past for finding frequent substructures that produce toxic effect [6].

In this paper, graph mining has been efficiently used for finding frequent substructures in datasets belonging to

carcinogenic compounds [10][11][12] and dataset of estrogen receptors [2][4][8] taken from DSSTox. A database of substructures identified as carcinogenic and frequently occurring in different chemical compound databases has been prepared. This database can act as a repository to be referred for testing toxicity in terms of carcinogenicity in a new compound to be tested on human subjects.

The paper has been divided into the following sections. Section 2 discusses the *gSpan* algorithm. Section 3 gives analysis of the results obtained by applying graph mining algorithm on different datasets taken from Distributed Structure-Searchable Toxicity.

## 2 Details of *gSpan* algorithm

In order to identify frequent substructures from the graph database, support of the identified candidate substructures is computed. Support of a given substructure is defined as the number of distinct (not isomorphic) substructures in a database of graphs. '*gSpan*' algorithm [28] starts by generating different DFS (Depth First Search) trees followed by identifying their DFS codes. For generating DFS tree from a given graph, the graph is traversed in depth first search [7] fashion. For a given graph, there can be more than one DFS trees. The concept of forward and backward edge is used by *gSpan* in DFS tree construction where forward edge set is defined as the set of edges from the graph that forms part of the DFS tree and backward edge set is the set of edges of the graph which are not in the DFS tree. For each DFS tree, DFS code is generated based on the lexicographical ordering defined on the vertex and edge labels. Among the different DFS code representations of a graph, the one with the minimum DFS code generated on the basis of lexicographic order is the *canonical representation* of the graph. A DFS tree is represented as a sequence of edges, called the DFS code, with each edge in the DFS tree being represented in terms of  $l_i$  and  $l_j$ , the labels of vertices  $v_i$  and  $v_j$  respectively;  $e_{i,j}$  is the edge label where  $i$  and  $j$  denote the vertex identification numbers. The structure of a chemical compound is expressed in terms of a graph with nodes of the graph representing the atoms and edges of the graph representing the bonds. Consider the graph given in Figure 1(a) for which there can be more than one DFS tree that are isomorphic to it and are given in Figure 1(b), 1(c) and 1(d). The lexicographical ordering among the different DFS codes is identified and finally a search tree based on this lexicographical ordering is constructed.

In the first stage, frequent 1-edge substructures (which represent two atoms in a chemical compounds) that satisfy minimum support criteria are identified. Minimum support is the support threshold defined for identifying frequently occurring substructures in the chemical compound database.

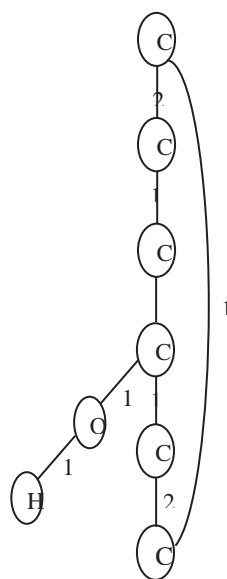


Fig. 1 (a)

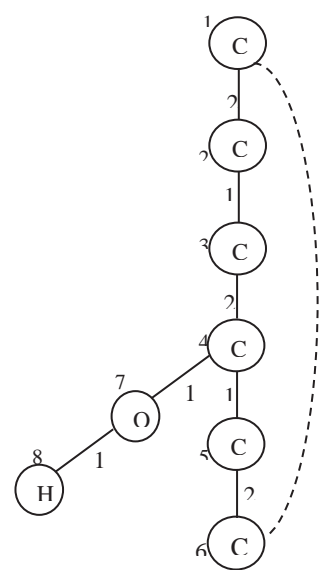


Fig. 1 (b)

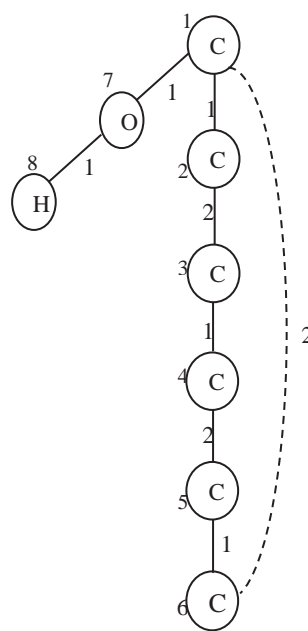


Fig. 1 (c)

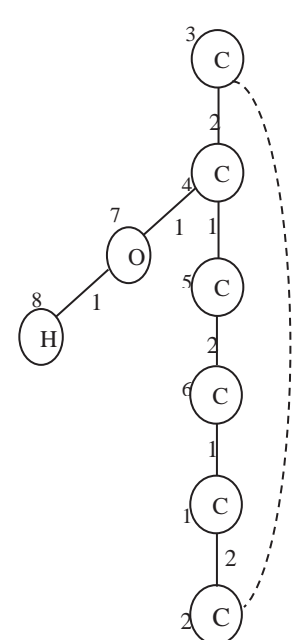


Fig. 1 (d)

If the support of a given substructure is greater than the minimum support, the substructure is identified as frequent. Using the concept of forward and backward extensions the substructures identified frequent at the previous level are extended to generate candidate substructures at the next level (see Fig.2 for chemical compound).



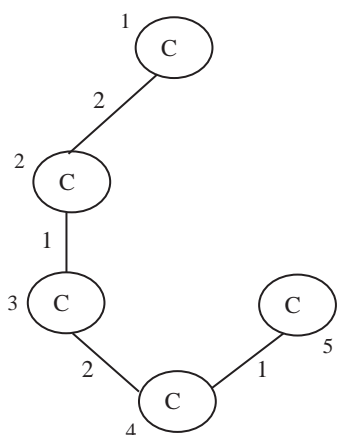


Fig. 2 (a) Frequent Substructure

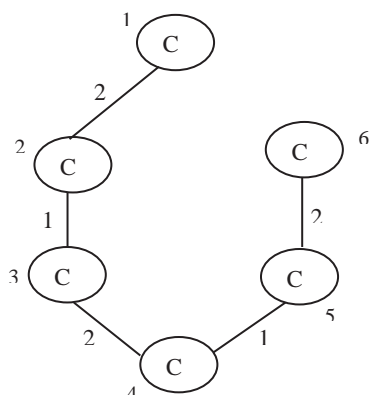


Fig. 2(b) Candidate Substructure generated using forward edge extension

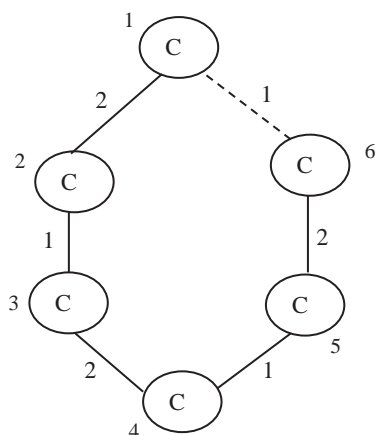


Fig. 2(c) Candidate Substructure generated from using backward edge extension

The process of recursively exploring substructures stops when either the support of the substructure is less than minimum support or its code is not a minimum code. 'gSpan' algorithm has been applied on the various datasets and relevant frequent substructures have been identified.

### 3 Results

The results obtained by applying graph mining algorithm on various datasets taken from Distributed Structure-Searchable Toxicity (DSSTox), a Public Database Network from the U.S. Environmental Protection Agency has been analyzed. The database uses a standard chemical structure annotation and is aimed at toxicology studies.

Frequent substructures related to toxicity have been identified using *gSpan* algorithm. All evaluations were carried out on Intel Pentium® 4 CPU 2.93GHz having 512 MB RAM. Results are compiled using g++ 3.0 compiler on Red Hat Linux 9 (2.4.20-8). Three datasets CPDBAS (Carcinogenic Potency Database Summary Tables - All Species) [10][11][12], NCTRER (FDA's National Center for Toxicological Research Estrogen Receptor Binding Database File) [2][4][8] and DBPCAN (EPA Water Disinfection By-Products with Carcinogenicity Estimates Database File [27] are used. Brief description of each of the datasets along with the relevant frequent substructures identified is shown in the following subsections.

#### (a) CPDBAS Dataset

The CPDB Summary, Tables list the summarized results for experiments on 1547 substances in the Carcinogenic Potency Database (CPDB). These Summary Tables report the strongest evidence of carcinogenicity for each chemical, in each sex/species and represent one of many possible summarizations of the data in the CPDB. Since, the substructures present in most of the chemical compounds are required; the minimum support is taken to be of a high value. By applying *gSpan* algorithm, frequent substructures obtained from this dataset at minimum support of 99.7% belongs to the family of propyl- and methyl- hydrazines, byphenyl and nitroso(propyl)amines. These frequent substructures are given in Figure 3. The toxic effect associated with the frequent substructures is discussed as follows:

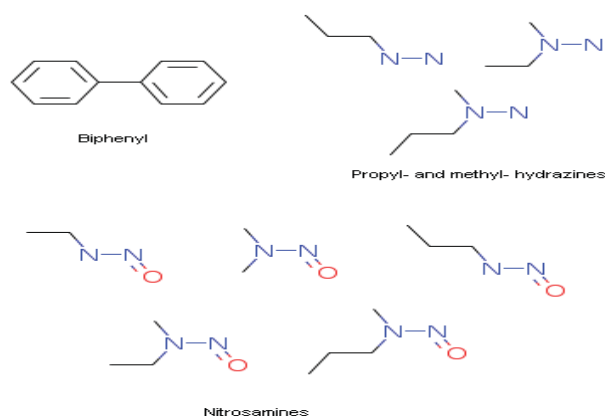
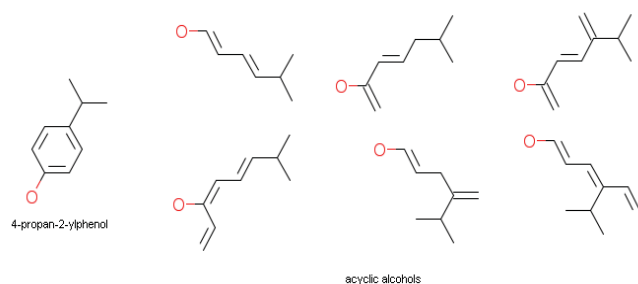


Figure 3 Relevant frequent substructures identified from CPDBAS dataset

Methylhydrazine is associated with many diseases, mainly cancer [17]. Inflammation of pancreas and infertility in males are some of the diseases related with PAPA-NONOate that contains propylhydrazine as its substructure [25]. Nitroso(propyl)amines and nitrosamines interact with cytochrome CYP2A6 and can cause diseases related to liver such as hepatitis B and C, cirrhosis, liver cancer and other types of cancer [13]. Diseases like tumor of liver and methemoglobinemia are associated with the decreased activity of hepatic cytochrome CYP1A2 enzyme, caused by biphenyl molecule [19]. Many biphenyl derivatives are associated with bioaccumulation [3], endocrine disruption [5] and neurotoxicity [24] because of their symmetry, hydrophobicity and ease of conjunction with halogen atoms. These identified substructures should be avoided in new drug design.

### (b) Dataset NCTREER

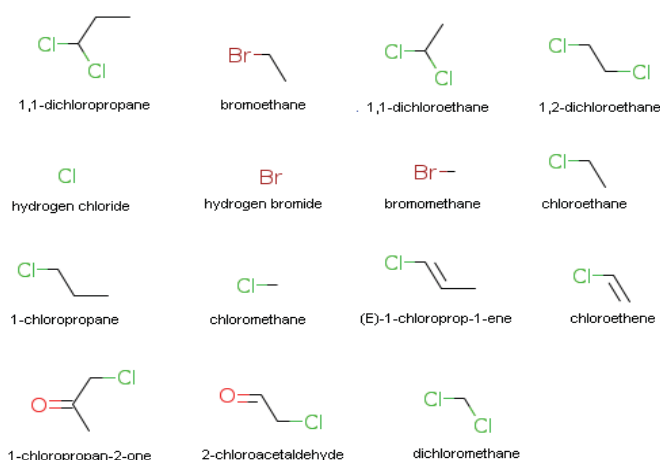
Database of experimental estrogen receptor (ER) binding results was generated by researchers within FDA's National Center for Toxicological Research (NCTR) for the purpose of developing improved QSAR model for predicting ER binding affinities. The NCTREER database provides activity classifications for a total of 232 chemical compounds selected *a priori* based on structural characteristics and tested in a well validated and standardized *in vitro* rat uterine cytosol ER competitive-binding assay [2] [4][8]. By applying *gSpan* algorithm frequent substructures obtained from this dataset at minimum support of .7% belongs to the family of acyclic alcohols and 4-(propan-2-yl) phenol. Frequent substructures of significance identified from NCTREER dataset are given in Figure 4. The frequent substructures identified from the NETREER dataset embrace well known hormonal effects, mainly due to their affinity for the estrogen receptors, for example, 17 beta estradiol (an active metabolic product of testosterone) matches with 4-(propan-2-yl) phenol substructure [22], diethylstilbestrol and bisphenols analogues on the other hand matches with acyclic alcohols. If the chemical compounds having these substructures do not show any carcinogenic effect, then, these chemical compounds can be further investigated for other properties and can be helpful in new drug design for removing estrogen deficiency.



**Figure 4** Relevant frequent substructures identified from NCTREER dataset

### (c) Dataset DBPCAN (EPA Water Disinfection By-Products with Carcinogenicity Estimates Database File)

The DBPCAN database consists largely of comparatively simple aliphatic organic structures, almost half halogenated, placed into one of 18 general chemical functional classes. The water disinfection by-products database contains calculated estimates of carcinogenic potential for 209 chemicals. The DBP's are grouped by a semiquantitative ranking scale of high (H), high-moderate (HM), moderate (M), low-moderate (LM), marginal (mar), and low (L) levels of concern for possible carcinogenicity. By applying *gSpan* algorithm number of frequent substructures obtained from this dataset at minimum support of 99.7% belongs to the family of halogens and haloalkanes. Frequent substructures identified from DBPCAN dataset are given in Figure 5.



**Figure 5** Frequent substructures identified from DBPCAN dataset

Almost all substructures obtained are associated with some kind of toxic effect. Bromomethane was used for pest control by farmers before it was related to diseases associated with people working in agricultural fields and heart diseases [21]. Hence, the chemical compounds with these substructures should not be used in drugs. A drug with the name Remoxipride was withdrawn from the market due to the presence of bromomethane substructure that leads to aplastic anemia [18]. Presence of bromine is not desired by therapeutic chemists of good repute [20].

If a repository of frequent substructures identified in different toxic compound databases can be made and preserved, it will be of great help in new drug design. When a new drug is identified, it is checked for the existence of these toxic substructures in it. If any of these structures exists in the chemical compound, it is not explored further on other aspects as it is not suitable for human beings.

## 4 Conclusions

The paper presents an application of graph mining algorithms for identifying frequent substructures from the database of chemical compounds. The analysis of the results on different datasets taken from Distributed Structure-Searchable Toxicity (DSSTox) has been done and presented. The scope of graph mining algorithms in identifying toxicity from chemical compounds has been investigated with the objective of reducing the search space being explored in drug discovery process.

## 5 References

- [1] R. Aggarwal and R. Srikant, "Fast algorithms for mining association rules", In Proc. Of 20<sup>th</sup> Intl. Conf. Very Large Data Bases (VLDB'94), pages 487-499, 1994.
- [2] R. M. Blair, H. Fang, W. S. Branham, B. S. Hass, S. L. Dial, C. L. Moland, W. Tong, L. Shi, R. Perkins and D. M. Sheehan, "The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands", Toxicol. Sci., Vol. 54, pages 138-153, 2000.
- [3] M. Van den Berg, L. Birnbaum, A. T. Bosveld, B. Brunstrom, P. Cook, M. Feeley, J. P. Giesy, A. Hanberg, R. Hasegawa, S. W. Kennedy, T. Kubiak, J. C. Larsen, F. X. van Leeuwen, A. K. Liem, C. Nolt, R. E. Peterson, L. Poellinger, S. Safe, D. Schrenk, D. Tillitt, M. Tysklind, M. Younes, F. Waern and T. Zacharewski, "Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife", Environ. Health Perspect., Vol. 106, Issue 12, pages 775-792, 1998.
- [4] W. S. Branham, S. L. Dial, C. L. Moland, B. S. Hass, R. M. Blair, H. Fang, L. Shi, W. Tong, R.G. Perkins and D.M. Sheehan, "Binding of phytoestrogens and mycoestrogens to the rat uterine estrogen receptor", J. Nutr., Vol. 132, pages 658-664, 2002
- [5] K. R. Chauhan, P. R. Kodavanti and J. D. McKinney, "Assessing the role of ortho-substitution on polychlorinated biphenyl binding to transthyretin, a thyroxine transport protein", Toxicol. Appl. Pharmacol., Vol. 162, Issue 1, pages 10-21, 2000.
- [6] R. N. Chittimoori, L. B. Holder and D. J. Cook, "Applying the *subdue* substructure discovery system to the chemical toxicity domain".
- [7] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms", MIT Press, 2001, Second Edition.
- [8] H. Fang, W. Tong, L.M. Shi, R. Blair, R. Perkins, W. Branham, B.S. Hass, Q. Xie, S.L. Dial, C.L. Moland and D.M. Sheehan, "Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens", Chem. Res. Tox. Vol. 14, pages 280-294, 2001.
- [9] J. Graham, C. D. Page and A. Kamal, "Accelerating the drug design process through parallel inductive logic programming data mining", CSB '03, page 400, 2003.
- [10] L. S. Gold, T. H. Slone, B. N. Ames, N. B. Manley, G. B. Garfinkel and L. Rohrbach, "Chapter 1: Carcinogenic Potency Database", In L. S. Gold and E. Zeiger, Eds. Handbook of Carcinogenic Potency and Genotoxicity Databases. Boca Raton, FL: CRC Press, pages 1-605, 1997.
- [11] L. S. Gold, N. B. Manley, T. H. Slone and J.M. Ward, "Compendium of chemical carcinogens by target organ: Results of chronic bioassays in rats, mice, hamsters, dogs and monkeys", Toxicol. Pathol., Vol. 29, pages 639-652, 2001.
- [12] L. S. Gold, N. B. Manley, T. H. Slone, L. Rohrbach and G.B. Garfinkel, "Supplement to the Carcinogenic Potency Database (CPDB): Results of animal bioassays published in the general literature through 1997 and by the National Toxicology Program in 1997 and 1998", Toxicol. Sci., Vol. 85, pages 747-800, 2005.
- [13] R. Grobholz, H. J. Hacker, B. Thorens and P. Bannasch, "Reduction in the expression of glucose transporter protein glut 2 in preneoplastic and neoplastic hepatic lesions and reexpression of glut 1 in late stages of hepatocarcinogenesis", Cancer Res., Vol. 53, pages 4204-4211, 1993.
- [14] J. Huan, W. Wang and J. Prins, "Efficient mining of frequent subgraphs in the presence of isomorphism", In Proc. of the 3<sup>rd</sup> IEEE Intl. Conf. on Data Mining, (ICDM'03), Piscataway, NJ, USA, IEEE Press, pages 549-552, 2003
- [15] L. B. Holder, D. J. Cook and S. Djoko, "Substructure discovery in the *subdue* system", In Proc. AAAI'94 Workshop Knowledge Discovery in Databases (KDD'94), pages 169-180, Seattle, WA, 1994
- [16] A. Inokuchi, T. Washio and H. Motoda, "An apriori-based algorithm for mining frequent substructures from graph data", In Proc. of 4<sup>th</sup> European Conf. on Principles and Practices of Knowledge Discovery in Databases, (PKDD'00), pages 13-23, 20 [17] Y. I. Kim, I. P. Pogribny, R. N. Salomon, S. W. Choi, D. E. Smith, S. J. James, and J. B. Mason, "Exon-specific dna hypomethylation of the p53 gene of rat colon induced by dimethylhydrazine modulation by dietary folate", Am. J. Pathol., Vol. 149, Issue 4, pages 1129-1137, 1996.
- [18] C. Köhler, H. Hall, O. Magnusson, T. Lewander and K. Gustafsson, "Biochemical pharmacology of the atypical neuroleptic remoxipride", Acta Psychiatr. Scand., Vol. 82, Issue S358, pages 27-36, 1990.
- [19] L. E. Korhonen, M. Rahnasto, N. J. Mähönen, C. Wittekindt, A. Poso, R. O. Juvonen, and H. Raunio, "Predictive three-dimensional quantitative structure-activity relationship of cytochrome p450 1a2 inhibitors", J. Med. Chem., Vol. 48, Issue 11, pages 3808-3815, 2005.
- [20] H. Kubinyi, "Computational and Structural Approaches to Drug Discovery", Cambridge: Royal Society of Chemistry, 2007.
- [21] P. K. Mill and R. Yang, "Prostate cancer risk in california farm workers", J. Occup. Environ. Med., Vol. 45, Issue 3, pages 249-258, 2003.
- [22] E. Nasatzky, Z. Schwartz, W. A. Soskolne, B. P. Brooks, D. D. Dean, B. D. Boyan and A. Ornoy, "Evidence for receptors specific for 17 beta-estradiol and testosterone in chondrocyte cultures", [Connect Tissue Res.](#), Vol. 30, Issue 4, pages 277-294, 1994.
- [23] D. Plewczynski, "Tvscreen: Trend vector virtual screening of large commercial compounds collections",

BIOTECHNO '08, June 29 – July 5, 2008, Bucharest, pages 59–63, 2008.

[24] T. Simon, J. K. Britt and R. C. James, “Development of a neurotoxic equivalence scheme of relative potency for assessing the risk of pcb mixtures”, *Regul. Toxicol. Pharmacol.*, Vol. 48, Issue 2, pages 148–170, 2007.

[25] A. C. Skinn and W. K. MacNaughton, “Nitric oxide inhibits camp-dependent cftr trafficking in intestinal epithelial cells”, *Am. J. Physiol. Gastrointest. Liver Physiol.*, Vol. 289, Issue 4, pages G739–G744, 2005.

[26] H. van de Waterbeemd and E. Gifford, “Admet *in silico* modelling: towards prediction paradise?”, *Nat Rev Drug Discov.*, Vol. 2, Issue 3, pages 192–204, March 2003.

[27] Y. T. Woo, D. Lai, J. L. McLain, M. K. Manibusan and V. Dellarco, “Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products”, *Environ. Health Perspect.*, 110 Suppl 1, pages 75-87, 2002.

[28] X. Yan and J. Han, “*gSpan*: Graph-based substructure pattern mining”, In *Proc. of ICDM*, pages 721-724, 2002.

## **SESSION**

# **INFORMATICS FOR EDUCATION AND RESEARCH IN HEALTH AND MEDICINE**

**Chair(s)**

**TBA**



# Application of Procedural Generation as a Medical Training Tool

J. Duffy<sup>2</sup> and Z. Wang<sup>1,2,\*</sup>

<sup>1</sup> College of Information Technology, Taiyuan University of Technology, Taiyuan, Shanxi, China

<sup>2</sup> Department of Math/CS, Virginia Wesleyan College, Norfolk, Virginia, USA

**Abstract:** *Procedural Generation is an algorithmic process that enables the creation of a variety of content without the need to create assets specific for each case. The purpose of this project is to explore a practical application procedural generation through the use of a basic technique of this process, a pseudo random number generator (PRNG), to create a simple medical training simulation. The simulation is able to create a variety of cases with unique details (such as ages and names), and has the potential to be scaled up to accommodate additional information or assets as needed. Resulting cases so far disassociate background information from the actual symptoms generated for a case, but with the implementation of an input system for a user's answers it serves as a functioning programming example of the procedural generation process.*

**Keywords:** Information systems, medical informatics, procedure generation

## 1. Introduction

“Procedural Generation” is a programming technique that allows for the creation of content through the use of an algorithm, rather than having to manually create it all beforehand. It is used in various media, some of the most common being computer graphics and electronic games. One approach to this method is to create the resources and assets for the program to choose from, and then having “assemble” content in a way specified by the programmer, which is the demonstrated method in this project. While procedural generation has gained an interest in computer graphics and games [1], there are much fewer mentions of it being used outside of those fields.

An interest in finding a practical use for procedural generation, as well as interest in the topic itself, led to the creation of this research project. This study explores the practical applications of procedural generation techniques. In particular, their use as a medical training tool [2]. By using a technique of procedural generation, the pseudorandom number generator, a basic medical diagnostic simulation was constructed in the Java programming language. The technique allows for the creation of many different cases without having to create a large amount of assets, reflecting the variation a medical professional would encounter in their line of work. The program is able to create a large variety of cases as well as prompt the user for diagnostic input. The program then checks to see if the input is correct based on the generated patient's case, and gives feedback accordingly.

The paper is organized as follows. The system overview is presented in the next section. The case study for most of the simulation programs is discussed in the section 3. A conclusion is followed in the last section.

## 2. System Overview

The goal of the project was to find a simple yet practical application of procedural generation. This was achieved by planning and creating a very basic (medically speaking) training program for those studying or practicing in the medical field.

**The PRNG:** The “PseudoRandom Number Generator,” (referred to as the “PRNG”) is one of the elementary procedural generation techniques, and was selected due to its relative ease of implementation in the project. They are called “pseudorandom” because they rely on something predictable in order to create a random number. For this project, the PRNG relies on the system's internal clock time in milliseconds to create the numbers. The program itself then formats

---

\* Corresponding author. Email: zwang@wvc.edu

those numbers into a form it uses to create a patient case.

**The Patient:** Given the complexity and the sheer amount of information that it used to create an actual, “real life” medical case and project time constraints, the patient cases presented by the program will have only the basic details. A generated case will contain a patient’s first and last name, blood type, and age, as well as basic medical symptoms that they are presenting.

**User Input:** The program generates a multiple choice prompt for the user based upon the details contained within the patient’s data. This method of user input was selected as it would be a familiar format for the user and its ease of use, in addition to it suiting the simplistic nature (medically speaking) of the cases.

### 3. Case Study

#### 3.1 PRNG.JAVA

The PRNG (PseudoRandom Number Generator) class uses the system’s hardware clock in milliseconds to create a random number, and is one of many procedural generation techniques. The PRNG class truncates the number it generates to four digits, which are used by the program itself to determine the characteristics of the Patient case. A summary of the PRNG code, including its methods and variables (excluding those used for testing) is as follows. An example is shown in Figure 1.

##### Variables

*Random pseudo* – A Random number generator that uses the method *System.currentTimeMillis()* to create a random double.

*int seeded* – a four element array that is used to store four digits of the generated number

##### Methods

double *prng()* – returns the double created by the number generator *pseudo*.

void *numGen()* – uses *prng()* to generate a number, converts that number to a string, creates a substring then converts that substring to integer digits that are stored in *seeded*.

int *getDigit(int e)* – returns the integer stored at element *e* in *seeded*.

int *getDubDigit(int a, int b)* – returns the digits stored in elements *a* and *b* in *seeded* as a double digit integer.

```
Our truncated random number: 1752
Our truncated random number: 3239
Our truncated random number: 4625
Our truncated random number: 0151
Our truncated random number: 4356
Our truncated random number: 2706
Our truncated random number: 0903
Our truncated random number: 0029
Our truncated random number: 4119
```

Figure 1 The numbers generated by the PRNG are truncated to 4 digits, which determine the characteristics of the patient

#### 3.2 SYMPTOM.JAVA

The Symptom data class is used in the construction of the Patient case, and is used as part of the user prompt in the Main method file. The actual Symptom object itself stores an ID number, its name and description as well as its solution. However, this solution is not used by the object but rather in the user prompt. The class contains several “get” and “set” methods used to respectively return and change the data contained within a Symptom object. However, most of these methods are almost identical in structure and were mostly used for debugging and testing, and as such will be excluded from the following summary. Additionally, only the full constructor is listed for similar reasons:

##### Variables:

int *id* – an integer that serves as the Symptom’s identification number

String *desc* – contains the Symptom’s description

String *name* – the Symptom’s name

String *ans* – stores the Symptom’s answer, which is used in the Main method

##### Constructor:

Symptom(int *i*, String *n*, String *d*, String *a*) – instantiates a Symptom object with the specified data. If no data is set, the program uses the default constructor.



Methods:

void *setData*(int *i*, String *n*, String *d*, String *a* ) – sets the data of a Symptom object with the specified data

String *toString*() – returns a String containing the name, ID number and description of the Symptom object, shown in Figure 2.

```
public String toString()
{ return name+" (SYMPTOM ID: "+id+"): "+desc;}
```

Figure 2 The *toString*() method illustrates the object's data structure

### 3.3 SYMPDAT.JAVA

SympDat, short for “Symptom Data,” is a class used primarily in the construction of the data matrix to be used in the Main method file (shown in Figure 3). The two constructors allow it to take either the same data used to instantiate two Symptom objects or two, actual Symptom objects. The SympDat class itself is not overly complex, and contains many of the same methods used in the Symptom class. The only difference is that there are redundant methods that only apply to a specific Symptom object stored within a SympDat object, but otherwise contains the same “get” and “set” methods as well as a *toString*() method that returns a String containing Symptom data in a similar fashion.

```
public SympDat(Symptom s1, Symptom s2)
{
    sy1 = s1;
    sy2 = s2;
}
```

Figure 3 A SympDat object contains two Symptom objects, for later use in the Main.java file

```
public Patient(String fN, String lN, String bT, int a, Symptom o, Symptom t, Symptom th, Symptom f)
{
    fName = fN;
    lName = lN;
    bType = bT;
    age = a;
    s1 = o;
    s2 = t;
    s3 = th;
    s4 = f;
}
```

Figure 4 The Patient Constructor

The Patient object takes three Strings, (first and last name, blood type) an int (age) and four Symptom objects to fully instantiate. The default constructor creates a Patient without any names or blood type, sets age to 20 and manually sets the Symptom ID numbers. The purpose for this object is to create the cases in

### 3.4 PATIENT.JAVA

The Patient class is one of the two major components to the program itself, as it contains the actual data itself used to create a Patient case and utilizes other classes to a large extent. It instantiates a PRNG object and then uses the numbers generated to determine which details are used to construct a Patient object. The data includes details such as names and blood type, with other variables being set through the use of the included methods.

Variables:

String *fName* – first name of a Patient

String *lName* – last name of a Patient

String *bType* – blood type of a Patient

int *age* – age of a Patient

Symptom *s1* through *s4* – the four Symptom variables store two Symptoms that are to be used in Main as correct answers and two that are used as wrong answers

Data:

String[] *firstName* – contains all of the first name Strings used by the program (contains 10 Strings).

String[] *lastName* – contains all of the last name Strings used by the program (contains 10 Strings).

String[] *bloodType* – contains all 8 medically known human blood types as Strings.

Constructor: See Figure 4.

Main and to demonstrate the procedural generation aspect of the program as a whole.

Methods:

void *createSymptoms()* – takes two digits given by PRNG, formats them into a form useable by the method, and then sets the ID numbers of two the first two Symptom objects. Then it sets the ID number of the Symptoms used for the wrong answers. Included within the method's programming are checks that ensure no two ID numbers are alike, ensuring that they are all unique and that there are no overlaps.

void *grabName()* – takes two digits given by PRNG and selects those respective elements from the *firstName* and *lastName* String arrays, then sets fName and lName.

void *genBlood()* – takes one digit given by PRNG and uses it to select a blood type from the *bloodType* array. Since there are only 8 medically accepted blood types, there is also a check to ensure that the method doesn't attempt to select an element outside of the array's index.

void *createPatient()* – a combination of *createSymptoms()*, *grabName()*, and *genBlood()*.

```
Scanner s = new Scanner(System.in);
String a = "y";
Boolean again = true;

//Our symptom data
Symptom s1 = new Symptom(0, "Ache", "Patient has aching muscles", "Advise mild pain medication
Symptom s2 = new Symptom(1, "Cold", "Patient is showing signs of a basic cold", "Advise increa
Symptom s3 = new Symptom(2, "Bruise", "Patient has visible bruising from a sport injury", "App
Symptom s4 = new Symptom(3, "Cut", "Patient has a visible, mildly bleeding cut", "Clean the af
Symptom s5 = new Symptom(4, "High Temperature", "Patient's body temp. is higher than average",

Symptom[] data = {s1, s2, s3, s4, s5}; //an array of our symptoms - to fill the data matrix ea
SympDat[][] sd = new SympDat[5][5]; //The data matrix

Patient test = new Patient(); //Our "patient"
```

Figure 5 Data and variables used in Main.java

The beginning code block (Figure 6) contains the data and variables used for the prompt later on in the program. Note the structure of a fully instantiated Symptom object contains its ID, a description and its solution.

In Figure 7, the program initializes the data for the SympData matrix. Once completed, the program then continues into the main code loop. It begins by setting the data for the Patient, and then setting the data for the Symptom objects contained within Patient with information from the matrix.

There is also a function to set the age of the Patient, achieved by using the PRNG's *getDubDigit()* method.

String *toString()* – this method is particularly important, as it is used by Main when outputting cases to the screen. It returns a formatted String containing the first and last name of a patient, the blood type, age, and the first two Symptom objects. It does not output the last two Symptom objects, as they are intended for use in Main to create the wrong answers.

### 3.5 MAIN.JAVA

This is the code that creates the cases, asks for user input, and then prompts for user input. There are no methods defined in this class, but instead contains fully instantiated Symptom data to be used by the cases themselves. It is able to create a new case for the user each time the program is run or loops. The code for Main.java is summarized in Figure 5.

```
//This loop initializes the data in the SympDat matrix
for(int c = 0; c<5; c++)
{
    for(int d = 0; d<5; d++)
    {
        if(d != c)
        {
            sd[c][d] = new SympDat(data[c], data[d]);
        }
    }
}

while(again) //A while loop in place in case the user wants to run
{
    test.createPatient(); //create the patient

    //set the data for the Symptom contained in the Patient object
    for(int z = 0; z < 5; z++)
    {
        int o = test.getSympOneID();
        int t = test.getSympTwoID();
        int th = test.getSympThreeID();
        int f = test.getSympFourID();

        if (o == data[z].getID())
            test.setSympOne(data[z]);
        else if (t == data[z].getID())
            test.setSympTwo(data[z]);
        else if (th == data[z].getID())
            test.setWrongOne(data[z]);
        else if (f == data[z].getID())
            test.setWrongTwo(data[z]);
    }
}
```

Figure 6 The initializing code

In Figure 8, the code creates a case and user prompt based entirely on the information contained within the Patient object. The prompt asks for two user integer inputs, which correspond to the numbers presented by the program in a “multiple choice” format. Note that the “*wrongOne()*” and “*wrongTwo()*” methods are the same as a “get” method but under a different name, hence why they were not mentioned earlier.

```

System.out.println(test.toString());

int correct1 = test.getSympOneID();
int correct2 = test.getSympTwoID();
int wrong1 = test.getSympThreeID();
int wrong2 = test.getSympFourID();

//The list of choices for the prompt
System.out.println("\nYour following treatment options: ");
System.out.println(test.getSympOneID()+" "+test.getAnsOne());
System.out.println(test.getSympTwoID()+" "+test.getAnsTwo());
System.out.println(test.getSympThreeID()+" "+test.wrongOne());
System.out.println(test.getSympFourID()+" "+test.wrongTwo());

//Prompt user to enter their answers
System.out.println("Enter the number corresponding to correct treatment: ");
int ans1 = s.nextInt();
System.out.println("Enter the second number corresponding to correct treatment: ");
int ans2 = s.nextInt();

```

Figure 7 Creating the user prompt

```

else if(ans2 == wrong1 || ans2 == wrong2)
{
    System.out.println("INCORRECT: Your second answer is wrong");

    if(ans1 == wrong1 || ans1 == wrong2)
        System.out.println("INCORRECT: Your first answer is also wrong");
}

//Ask if they would like to do another case
System.out.println("Try another case? (y/n)");
a = s.next();

//Check to see if they said "y"es
if(a.equalsIgnoreCase("y"))
    again = true;
else
    again = false;

}

```

Figure 8 A sample of the logic coding and a prompt asking if the user wants to run the program again

Finally, the program performs some logic checks to see if the user inputs were correct. It should be noted that not all of the logic coding is presented in the above picture, but the program is able to determine whether the user inputs are correct or incorrect and even specify which answers are incorrect. It then asks the user if they would like to run the code again and try another case, ending if they do not.

```

PATIENT NAME: Smith, John
BLOOD TYPE: O+
AGE: 34
SYMPTOMS:
  Cut (SYMPTOM ID: 3): Patient has a visible, mildly bleeding cut
  Cold (SYMPTOM ID: 1): Patient is showing signs of a basic cold

Your following treatment options:
3: Clean the affected area and apply a bandage
1: Advise increased fluid intake and rest
4: Advise rest and a mild fever reducer
2: Apply ice the affected area
Enter the number corresponding to correct treatment:
3
Enter the second number corresponding to correct treatment:
2
INCORRECT: Your second answer is wrong
Try another case? (y/n)

```

Figure 9 Example of demonstration displayed

Figure 9 demonstrates what the end user sees when they run the program, and what a fully generated case and prompt look like. For this particular case, our patient's full name is John Smith, aged 34, and has a O+ blood type. His symptoms include a bleeding cut and he appears to have a cold. The treatment options are based upon the Symptom objects contained within the Patient object, hence why it stores four but only displays two of them, as the other two are used to create possible incorrect answers.

## 4. Conclusion

Through the creation of a small amount of assets, the program is able to create a large amount of cases through the use of algorithms, demonstrating its practical use in a training scenario where such a number of cases can occur in the field. None of the cases are manually created beforehand, and if they were, there would need to be 2.4 million premade cases – a task that would take a large amount of time through traditional methods and create a large data

file. The data used in the project can be easily increased to allow for many more cases to be generated by the program. Furthermore, additions to the program can increase its worth as a training tool, such as including the patient's temperature or having the patient's medical details influence what symptoms they present, and would be possible with more time and testing.

## 5. Reference

- [1] H. Smith. Procedural Trees and Procedural Fire in a Virtual World, White Paper, Intel Software and Services Group, 2008.  
<https://software.intel.com/sites/default/files/Procedural-Trees-and-Procedural-Fire-Wp.pdf>
- [2] E. H. Shortliffe, et al. Medical Informatics, Computer Application in Health Care and Biomedicine. 2<sup>nd</sup> ed. Springer-Verlag. 2001.

# Socially Mediated Health Information for Dengue Surveillance and Education in Sri Lanka

Schubert Foo, May O. Lwin and Vajira S. Rathnayake

<sup>1</sup>Centre of Social Media Innovations for Communities (COSMIC), Nanyang Technological University, Singapore

**Abstract** – *Dengue affects nearly 50 million people annually and specially threatens populations living in tropical regions like Sri Lanka. Based on the findings of a needs assessment of dengue management practices in Colombo, Sri Lanka, we have designed a socially mediated health information system for dengue surveillance and education. This system integrates the power of digital surveillance, digitized dengue monitoring and mapping, and digitized dengue education. We identify specific loopholes in existing workflows and describe how our system has been designed to address these, resulting in greater efficiency leading to better dengue management. The paper culminates with a discussions of lessons learned from our experiences and the next steps of our applied innovation.*

**Keywords:** dengue; social media; surveillance; education

## 1 Background

With nearly 50 million cases occurring annually, dengue poses an immense threat to global health, especially populations living in tropical regions [1]. Sri Lanka, a small island nation located in the Indian subcontinent, exemplifies this situation with dengue epidemics over the period 2010-2015 recording more than 200,000 cases [2]. The epicenter of the dengue epidemic in Sri Lanka curiously lies in the Western Province – home to the nation's political capital of Colombo – that alone accounts for nearly half of the country's dengue burden [3].

Sri Lanka's challenges in grappling with dengue since 2009 have interestingly coincided with the meteoric rise of mobile penetration rates in the aftermath of the civil war. Mobile services in Sri Lanka are now among the most affordable in the world and have been increasingly used in the mobile health and education sectors. However the dengue information infrastructure, whose efficiency had been adversely affected by traditional paper-based systems, had failed to benefit from such technological interventions.

### 1.1 Technological interventions for dengue prevention and management

Elsewhere, technological interventions for bolstering dengue surveillance have mainly focused on the use of geographical information systems (GIS) and other surveillance systems to facilitate early notification or

warnings of potential outbreaks. For instance, Chang *et al* [4] used Google Earth and ArcGIS 9 to create a surveillance system in Nicaragua that can allow public health workers to identify high indices of mosquito infestation in relation to larval development sites like garbage piles and stagnant water pools. In Brazil, researchers developed the SMCP-Aedes, an entomological surveillance system focused on collecting, storing, analyzing and disseminating mosquito-related information online [5]. In Thailand, Ditsuwan *et al* [6] used a combination of a national surveillance system database and GIS to evaluate the burden of dengue and chikungunya fever. Dengue-GIS has also been used for monitoring and evaluating national-level epidemiological, entomological and control interventions in Mexico, and has found to be useful for decision-making at different levels of the dengue control system [7]. While these initiatives have attempted to use GIS for different aspects of dengue prevention and control programs in their respective countries, we recognize three main limitations in extant work. First, we notice a paucity of technological interventions that reach beyond the offices of health policymakers and authorities to influence the actual workflow of health workers at points where they interact with the public. Second, it is evident that most technological interventions are focused mainly on surveillance but rarely facilitate efficient health education or community engagement, which form some of the bedrocks of any powerful dengue prevention program [8][9]. Third, the cost-effective nature and multi-functional capabilities of mobile phones have been employed for a range of public health concerns in developing countries [10], but not as much for dengue.

## 2 Methodology

Drawing upon existing evidence and driven by the mandate of the Centre of Social Media Innovations for Communities' (COSMIC), we first conducted a field study to identify the core needs of public health inspectors (PHIs) in Sri Lanka in relation to dengue surveillance and prevention practices. The objective was to also identify areas of concern in existing practices that could be strengthened through the use of mobile-based technologies and social media.

We focused on the PHIs as potential adopters, as they constitute the last mile of the public health delivery system in Sri Lanka and are hence the most critical cog in the wheel of dengue prevention efforts. Recent reports reveal that existing inefficiencies in the public health system might be partially explained by the PHI's overburdening workload: one PHI in

Colombo covers the public health concerns of nearly 50,000 people. Colombo comprises 47 wards (administrative units) that are covered by six Medical Officers for Health (MOH) units. Each MOH is served by 5-7 PHIs. Our needs assessment was conducted through 30 in-depth interviews lasting 45-60 minutes in the premises of the Colombo Municipal Council (CMC).

### 3 Findings

We present three main summary findings that pertain to dengue information issues among the PHIs in Colombo.

#### 3.1 Current dengue information flow

We obtained a comprehensive understanding of the flow of dengue-related information between different agencies involved in the dengue surveillance programs in Colombo. As can be seen in Figure 1, the existing dengue information architecture reflects a circuitous and time-consuming process. This process commences with a patient who experiences symptoms visiting the hospital who in turn hand over a paper-based record of suspected dengue cases to the PHI who is assigned to that particular hospital. All PHIs who receive this information pass it along to the CMC's Epidemiological Unit,

where an official is assigned to create a separate file for individual patients. The official categorizes all these files according to the MOH jurisdiction under which they are covered, and dispatches this information to each of the MOH offices through the CMC's Epidemiological Unit (CMC-EU). The MOHs then distribute the files to their PHIs for follow-up through patient visits. Each PHI visits the patient to confirm his/her diagnosis for dengue, upon which a decision is taken to either fill a Communicable Disease Form (CDF) and a Dengue Investigation Form (DIF) or only the former, depending on whether the patient is tested positive or negative. Also, in case of a positive diagnosis, the PHI is required to conduct a house and area inspection to identify possible mosquito breeding sites and educate the patient and his family on protecting themselves from dengue. After obtaining the entire set of CDFs and DIFs from the PHIs under their jurisdiction, the MOHs officially approve the forms before dispatching them to the CMC-EU. The CMC-EU manually collates the information from all the DIFs to create a record, map dengue cases on a manual map, and ensure that all the cases are within the CMC jurisdiction. At the end of this process, a formal report is sent to the CMC's Public Health Department (CMC-PHD) who officially sign off on it before dispatching it to the Ministry of Health. The whole process could take anywhere from 7-10 days.

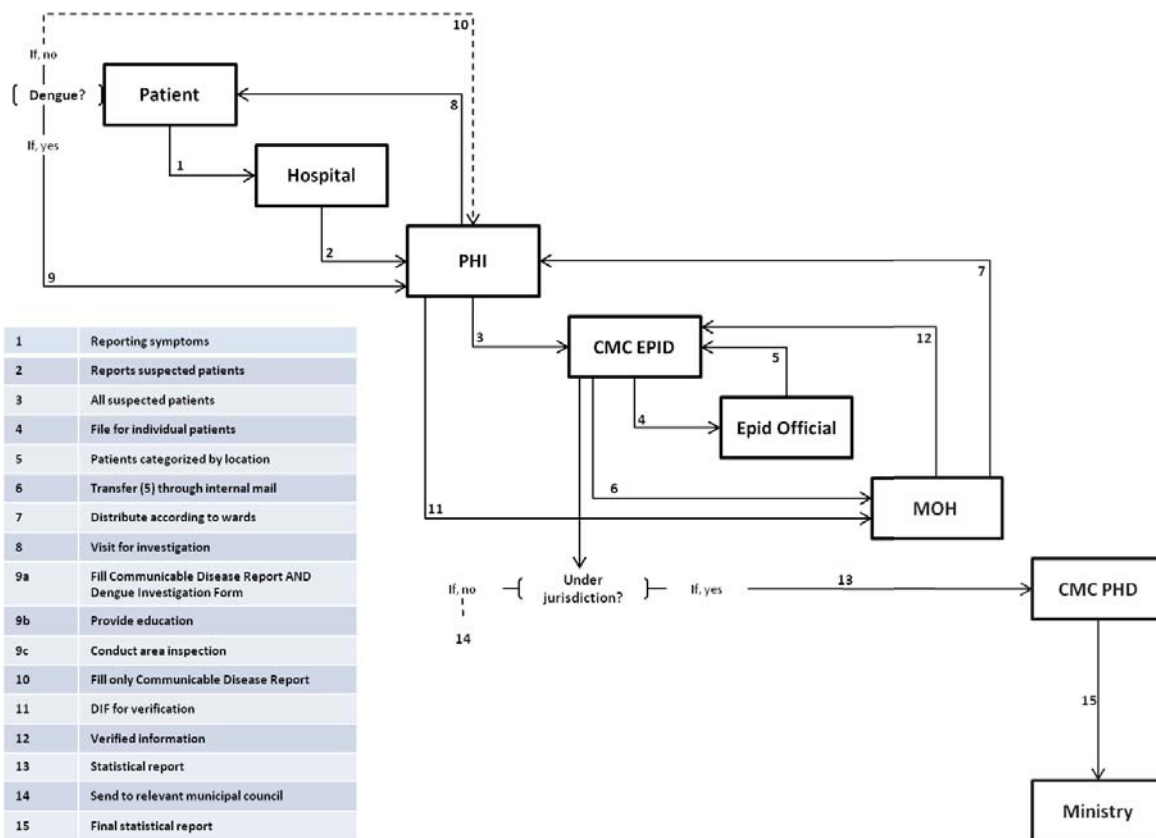


Figure 1: Typical flow of dengue information by the Colombo Municipal Council

### 3.2 Barriers in existing data collection and reporting practices

The workflow shown in Figure 1 suffers from a principal weakness in steps 8 and 9, where the PHI physically interacts with the client and obtains critical details about the case. The first challenge was recording information in the CDF and DIF which, the PHIs claim could consume between 30-60 minutes including being exposed to the risks of information inaccuracy and the work was itself logistically inconvenient. Secondly, the standalone GPS device used by PHIs was technically unreliable (lack of signals, inaccurate readings) and other officials found it difficult to decode the geographically coordinate quickly and efficiently. Lastly, PHIs were educating dengue-affected individuals and families using outmoded means of health communication such as pamphlets and brochures. With minimum persuasive impact and lack of audience engagement, PHI reported that these materials bore minimum effects on the attitudes, knowledge and behaviors of the public related to dengue.

### 3.3 Need for technological intervention

The profusion of media penetration in the aftermath of 2009 was also reflected in the number of PHIs who used simple mobile phones or smartphones. The younger PHIs were familiar with mobile applications and demonstrated enthusiasm for integrating mobile-based innovations into their workflow; the senior PHIs, despite their lack of prior experience with smartphones, recognized the need for using mobile devices as part of their daily tasks. Broadly, the PHIs expressed a need for a portable solution that would allow them to accurately record geographical coordinates, capture pictures and was powered by the Internet. These functionalities would ease their effort in executing the most challenging tasks related to dengue surveillance on a daily basis.

## 4 Mo-Buzz: A socially mediated system for dengue surveillance, engagement and education

The needs assessment helped to identify the key gaps and constraints in the existing dengue information flow and also opportunities to address these using mobile social media. The challenge was to facilitate easier and more efficient exchange of information between actors without changing the existing workflow that has been established according to national guidelines. Instead of digitally transforming the information flow in its entirety, the priority for our innovation would be to address the bottlenecks in steps 8 and 9 identified in Figure 1 and to facilitate a more effective and efficient client visit by the PHIs. The following sections describe our proposed solution, namely Mo-Buzz, which is a socially mediated system that is built upon integrated information architecture and is available for PHIs on portable tablets.

### 4.1 Technical specifications

The overall system architecture is shown in Figure 2. The system is built on open source technologies and is mainly purposed for mobile and web based application which can be accessed through an Android platform or a web browser. The Android solution forms part of the main application by running as an agent on mobile devices. The PHIs and MOH can report information in various forms (photo or text) using mobile devices. The web-based solution is designed mainly for the management as it offers an interactive system for geospatial visualization, reports for reported Dengue Investigation Form, summaries and graphs, and web forms for other details. The solution is developed using Java related technologies. The server side of this system is supported by Apache, Tomcat, and MySQL.





Mo-Buzz system offers a tablet-based health education component. This has been done in the backdrop of mounting evidence which suggests positive outcomes resulting from mobile-mediated health education modules for health workers in other contexts. The first version of the health educational

module includes digitized versions of the CMC's dengue education materials that the PHI presents to his clients complemented by verbal explanations of dengue prevention concepts. For future versions, we are building graphics, animations and tailoring capabilities into the health education component.

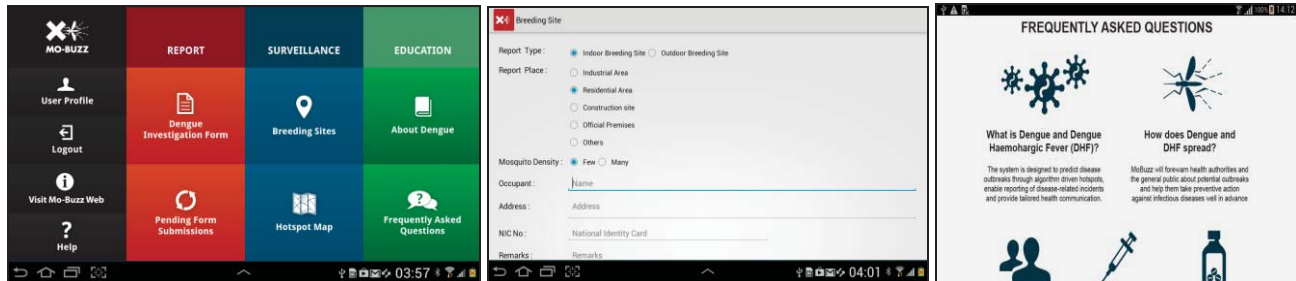


Figure 3: Screenshots from Mo-Buzz depicting the home screen (left), mosquito reporting form (center) and health educational component (right)

## 5 Current status and next steps

The Mo-Buzz system is currently being used by all 55 PHIs in the CMC. It is made available to them on tablets that are powered by a data plan from Mobitel, Sri Lanka's second largest telecommunication provider. In the future, the research team plans to integrate the PHI system with the Mo-Buzz dengue system for the general public. The general public version uses the concept of crowdsource-based disease surveillance and enables users to report mosquito breeding sites to the CMC through geo-tagged pictures captured with their smartphones. In addition, it also includes real-time dengue maps and dynamic health educational components that together provide necessary information for users to take adequate personal protection measures against dengue in advance, as they are constantly updated about dengue danger zones in their city.

## 6 Discussion

The introduction and adoption of Mo-Buzz among PHIs in Colombo provided several lessons from the standpoint of health information system innovations in developing countries. For instance, our conversations with PHIs and subsequent meetings revealed that the PHIs and other actors were primed to the existing dengue information system (despite its circuitous nature) and that a complete overhaul would cause important delays apart from causing systemic disturbances. We suspect that the organizational acceptance of Mo-Buzz might have been facilitated by a perception of our system as one that was seeking to make an incremental change, instead of a wide-ranging transformation. Secondly, we acknowledge the importance of local public-private partnerships in executing a project of this nature as common cultural attributes act as a catalyst in creating sustainable relationships. Lastly, we observe that the discourse surrounding health information technology interventions in developing countries is situated in the debate on ground-up

versus top-down approaches. Our intervention demonstrates that an integration of the two approaches is possible and might provide unique benefits. For instance, our "top-down" approach of using mobile social media functionalities (driven by the intellectual expertise available in our team) blended with a grounded understanding of the local needs and challenges, providing potential users with new options that might or might not have previously envisaged.

## 7 Conclusion

We have proposed and outlined the Mo-Buzz solution to overcome the weaknesses of the existing workflow and situation of managing the dengue situation in Colombo. The system for the PHIs is currently being used and evaluated along with the parallel development of the public Mo-Buzz version.

## 8 Acknowledgement

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

## 9 References

- [1] Farrar J, Focks D, Gubler D, et al. "Editorial: Towards A Global Dengue Research Agenda". *Tropical Medicine & International Health*, Vol. 12, No. 6, pp. 695-699, 2007.
- [2] Epidemiology Unit Ministry of Health: Distribution of Notification (H399) Disease Cases by Month. [http://www.epid.gov.lk/web/index.php?option=com\\_casesand\\_deaths&Itemid=448&lang=en#](http://www.epid.gov.lk/web/index.php?option=com_casesand_deaths&Itemid=448&lang=en#).

[3] Ministry of Defense: Troops to Be Deployed for Dengue Prevention.  
[http://www.defence.lk/new.asp?fname=Troops\\_to\\_deploy\\_for\\_dengue\\_prevention\\_20130617\\_02](http://www.defence.lk/new.asp?fname=Troops_to_deploy_for_dengue_prevention_20130617_02)

[4] Chang AY, Parrales ME, Jimenez J, et al. "Combining Google Earth and GIS Mapping Technologies in A Dengue Surveillance System for Developing Countries." *International Journal of Health Geographics*, Vol. 8, No. 1, 49, 2009.

[5] Regis L, Souza WV, Furtado AF, et al. "An Entomological Surveillance System Based on Open Spatial Information for Participative Dengue Control. *Anais da Academia Brasileira de Ciências*, Vol. 8, No. 4, 655-662, 2009.

[6] Ditsuwan T, Liabsuetrakul T, Chongsuvivatwong V, Thammapalo S, McNeil E. "Assessing the Spreading Patterns of Dengue Infection and Chikungunya Fever Outbreaks in Lower Southern Thailand Using A Geographic Information System. *Annals of Epidemiology*, Vol. 21, No. 4, 253-261, 2011.

[7] Hernández-Ávila JE, Rodríguez M-H, Santos-Luna R, et al. "Nation-wide, Web-based, Geographic Information System for the Integrated Surveillance and Control of Dengue Fever in Mexico. *PLoS One*, Vol. 8, No. 8, e70231, 2013.

[8] Heintze C, Garrido MV, Kroeger A. "What Do Community-based Dengue Control Programmes Achieve? A Systematic Review of Published Evaluations. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol. 101, No. 4, 317-325, 2007.

[9] McNaughton D. "The Importance of Long-term Social Research in Enabling Participation and Developing Engagement Strategies for New Dengue Control Technologies. *PLoS Neglected Tropical Diseases*, Vol. 6, No. 8, e1785, 2012.

[10] Curioso WH, Mechael PN. "Enhancing 'm-health' with South-to-south Collaborations. *Health Affairs*, Vol. 29, No. 2, 264-267, 2010.

**SESSION**  
**POSTER PAPERS**

**Chair(s)**

**TBA**



# Legislative Bias in Aggregating Electronic Health Records

Edward Brown,

Department of Computer Science, Memorial University, St. John's, NL, Canada

**Abstract** - *An examination of the developing legislative regime concerning digital health information reveals that the data repositories for Electronic Health Records favor large monolithic datasets over distributed information systems. This paper inquires into the practical and privacy-driven motivation behind such designs in the Canadian health care context and suggests alternatives which could be explored for their potential to improve information access.*

**Keywords:** Electronic Health Record, Access Control, Cryptography, Legislation, Privacy

## 1 Electronic Health Records

The Canadian health care system is largely public funded since its rapid development in the 1940's. Constitutionally a matter of provincial jurisdiction, the federal government became involved in health services through its spending power authority. By 1984, the Canada Health Act [1] normalized the Canadian federal government's involvement in funding health care, establishing criteria on portability, accessibility, universality, comprehensiveness, and public administration for access to federal funds. Interjurisdictional agreements on health care reform and standards are now encouraged with consolidated inter-governmental block funding transfer agreements. [2][3]

With a universal publicly funded system, and universal coverage as a long standing principle, a natural expectation is that the implementation of electronic health records would be a simple matter of putting Canadian health records on line, so they could be accessed from any location. In fact, such a vision of a pan-Canadian health record was promoted in a major federal project initiative starting with a budget allocation of 50\$ million to do exactly that in 1997 [4], and leading to expenditures of over 4\$ billion over the intervening years since. This grew into a concept of a seamless, boarderless information system free from red tape and wasteful duplication. [5] However, the development of such a system has not been a strait-forward progression.

The political and commercial barriers to the development of a universal Electronic Health Record would seem to be minimal in a jurisdiction with universal socialized medical care. In pre-Obamacare United States, the barriers related to sharing records had to do with insurance and payment information; not only is health information of commercial value to a significant insurance market, but particular health information affecting insurability and pre-existing conditions had paramount consequences for availability of affordable treatment. The rules for sharing of personal health information represented in HIPAA are substantially a response

to this market environment. [6] In the Canadian context, availability of care is not significantly affected by insurance and payee structure, and the rules for sharing health information evolved from a somewhat different set of considerations.

In particular, interoperability of systems was mandated early on in national efforts by the organizations that emerged to deal with digital infrastructure (such as Health Canada Infoway at the national level). Security of transmitted data was promoted by advocating specific software architecture standards. [7] Health information was to be shared among those involved in the treatment of the patient. However, this system also had to respond to a rapidly evolving legislative regime concerning digital privacy, entrenched by 2001 in sweeping new national privacy legislation [8] and popularized by a number of high profile and large scale privacy breaches of health information. [9]

Regulations and legislation soon responded to the privacy requirements in various Canadian jurisdictions surrounding health information. Partly mirroring HIPAA, but more prescriptive in character, these regulations created responsibilities for the maintenance and protection of personal health information by traditional health care workers. These workers become "information custodians" under these new laws, with correspondingly new obligations and liability shields for compliance with the new standards for custodianship of PHI (Personal Health Information). Information related to specific patients can be shared among "information custodians" whom are involved in primary care of the patient. [10]

## 2 Health Information Custodians

The thesis of this paper is that the construction of a custodial relationship to electronic records supports a bias towards aggregate collections of data for EHR implementations. Two cultural influences encourage a centralized custodial posture respecting EHRs:

(1) Clinical and other health service providers are traditionally used to having control over their patient records. This is reflected in a common (although legally questionable) attitude often recited that "we own the records, while the patient owns the information". There are a number of legitimate reasons why a health service provider wants to control their records. The records are necessarily for administration of their practice; they have to have confidence in the nature of the record contents; the notes and entries on a clinical chart may be particular in

style and purpose to an individual practice; patients may object to record entries for idiosyncratic and personal reasons, and may react strongly to certain types of data entries. Fundamentally, health care workers want to deal with records maintained by health care workers, not by patients. This leads naturally to the “custodial” posture reflected in legislation and physical control of records “residing” somewhere geographically. While these elements (control, content and physical custody) are relatively inseparable for paper records, they are independent in electronic records: the records need not be localized or in the physical custody of particular individuals to maintain valid and reliable content.

(2) Technological requirements of the emerging privacy regime are easier to impose as a centralized service. As an example, health privacy regulations have evolved in Canada to require role based access control (RBAC). [11] Thematically, the notion of “access control” is itself predicated on the concept of a data repository to be accessed. Furthermore, an authentication system based on “roles” mean the users have to be categorized and authenticated with role information maintained as data records – implying some kind of centralized service or data collection. Regulations now mandate that this particular technical mechanism is used for securing information.

There are alternatives to aggregating large data sets that satisfy a custodial paradigm. Public key infrastructure mechanisms have been mature enough for more than a decade to allow on-demand and multi-level decryption of secure records and data services. Rather than securing the data records themselves, it would be simpler to leverage off of existing authentication mechanisms that underly secure communication on the Internet. Information access could be time limited and security mechanisms revised as dictated by future technological advances. Regulations that required data records to be released according to a security encryption standard, would mean the records themselves would no longer have to be placed in a custodial centralized structure. Authorization to examine elements of a health record could be represented with an appropriate key.

### 3 Alternatives

The advantage to such a cryptography-based (instead of custodial-based) system is that the encrypted records could be freely distributed without worrying about unauthorized access to sensitive health information. The impetus towards large aggregated data sets would be eliminated. Instead of being constrained to particular technologies, innovations in information systems could be applied to Health information systems as they are developed. As examples to such alternatives, In the 1990's, a system predicated on distributed consent rather than securing records was implemented in Australia. [12] A mechanism where authorization policy was implemented as constraints on messaging instead of on record maintenance was illustrated in [13].

The custodial attitude towards Electronic Health Records is unlikely admit any change: it is embedded in legislation and currently deployed in the accepted digital health infrastructure. However, there is a context in which we can foresee a properly distributed health record system arising: personal health records (PHRs) maintained by individuals are becoming more popular, and notionally are as diverse as exercise and nutritional apps. [14] As these personal health tools become more popular and comprehensive, individuals may eventually expect to download EHR information into the PHR. The PHR could become the synchronization and verification tool for personal health information. Secure encrypted communication protocols would be an inevitable development.

### 4 References

- [1] *Canada Health Act*, Revised Statutes of Canada, 1985, c-6.
- [2] *Canada Health Transfer*, Department of Finance, Government of Canada, online <http://www.fin.gc.ca/fedprov/mtp-eng.asp>, accessed Jan 12, 2015.
- [3] *Accord on Health Care Renewal*, Health Canada, online <http://www.hc-sc.gc.ca/hcs-sss/delivery-prestation/fptcollab/2003accord/index-eng.php>, accessed Jan 12, 2015.
- [4] P. Martin, 1997 *Budget in Brief*, Department of Finance, Government of Canada.
- [5] A. Morris, 2005, *The electronic health record in Canada: the first steps*, Health Law Review 14(2).
- [6] *Health Insurance Portability and Accountability Act*, 1996, 104<sup>th</sup> U.S. Congress 191, Statutes at Large 110/136.
- [7] *Privacy and Security Requirements*, 2014, Health Canada Infoway, online <https://www.infoway-inforoute.ca/en/ressource-centre/advanced-search?q=architecture>, accessed Jan 24, 2015.
- [8] *Personal Information Protection and Electronic Documents Act*, Statues of Canada, 2000, c. 5.
- [9] *Government fears privacy breach in computer theft*, CBC News, Jan, 2003.
- [10] *Personal Health Information Protection Act*, Statues of Ontario, Canada, 2004, c. 3.
- [11] *Personal Health Information Privacy Policy*, eHealth Ontario, 2014, online: [http://www.ehealthontario.on.ca/images/uploads/pages/documents/PHI\\_PrivacyPolicy\\_EN.pdf](http://www.ehealthontario.on.ca/images/uploads/pages/documents/PHI_PrivacyPolicy_EN.pdf), accessed Feb 7, 2015.
- [12] O'Keefe, C.M., Greenfield, P., and Goodchild, A, 2005. *A Decentralised Approach to Electronic Consent and Health Information Access Control*. Journal of Research and Practice in Information Technology, 37(2).
- [13] Brown, E. and Goodyear, J., 2011. *Dynamic Routing Using Health Information*. Biomedical Eng, Systems, 127.
- [14] Sherer, S., 2014, *Patients are not Simply Health IT Users*, Communications of the Assoc for Information Systems, 34(17).

# A Semantic Approach for Predicting Symptomless Unprovoked VTE through Analysis of Risk Factors

Susan Sabra<sup>1</sup>, Khalid Mahmood<sup>1</sup>

<sup>1</sup>Computer Science and Engineering Department, Oakland University  
2200 N Squirrel Rd, Rochester, MI 48309 USA

*Abstract - This paper presents a novel approach using semantics to analyze clinical notes from electronic medical records (EMR) and wearable body sensors (WBS) data to predict the first occurrence of symptomless unprovoked venous thromboembolism (VTE) in patients. Our semantic approach extracts hidden factors from unstructured text that might be very critical in the diagnosis or identification of potential risks of having a VTE. Our scoring model for risk factors is used to calculate the clinical probability for a patient of having or developing a case of VTE. Our predictive analytics model consists of a supervised learning classifier using the calculated clinical probability and patients' record of risk factors to predict a patient potential risk probability.*

**Keywords:** Semantic, predictive, symptomless, unprovoked VTE, IoT.

## 1 Introduction

The goal of our work is the attempt to save lives in cases of symptomless unprovoked VTE that could be fatal from first occurrence. VTE comprises deep vein thrombosis (DVT), pulmonary embolism (PE), or both. It is the third common cardiovascular disorder [7] that affects people from both genders at an age as young as 20-years old. VTE can be provoked or unprovoked: provoked cases can occur usually after certain types of surgeries, unprovoked cases can occur for the first time without any symptoms and can be fatal. Venous ultrasonography and venography are the main currently used diagnosis techniques for detecting and diagnosing VTE [6]. Fatality associated with recurrent VTE remain unacceptably high. To monitor and evaluate a patient's health status, lifestyle data is obtained from wearable body devices and medical (personal and familial) data from the patient's EMR. Some WBS provide direct data such as heart rate and blood pressure while other data can be calculated based on the sensor's data such as BMI (body mass index) as an indicator for obesity. The structured data in an EMR is typically values for vital signs, height, weight, and blood test

results. The unstructured data may contain valuable personal information about the lifestyle of the patient and their familial health history. In clinical domain, much of the available clinical data is recorded as free-text [1-4].

Our model extracts structured and unstructured data from an EMR and a WBS in order to perform a semantic analysis to properly classify the risk factors keywords using the YTEX and NLP classifiers. Keywords are scored in order to calculate the clinical probability for patients of 20 years or older for developing a first-time symptomless unprovoked VTE. Our main contribution is in the originality of our approach to use the semantic analysis of unstructured clinical notes with medical ontologies. This approach properly classifies keywords and features of risk factors from EMR and WBS to calculate the clinical probability of VTE.

## 2 Related Work

The literature presents numerous studies focusing on the risk of recurrence of VTE, but no study has considered the prediction of the first lifetime occurrence of VTE in patients using a semantic approach. In [14] they focused on improving the prediction of the recurrence risk of VTE using a simple risk assessment model. It is evident from the numerous studies in the literature that no risk factor is an independent predictor for the occurrence of DVT or PE. A study published in 2000, [13] stated that in the US, 201000 first lifetime cases of VTE will occur where 25% of these cases will die within 7 days of the occurrence. Moreover, for 22% of all cases death is so rapid that there is insufficient time for intervention to save the life of the patient. In another study that focused on one risk factor: the long-haul flights [10], they found that 10% of the passengers developed symptomless unprovoked DVT in the calf regardless of the gender. A Norwegian study, published in 2012, found that the incident of all first events of occurrence of VTE was 1.43 per person-year but the risk was higher for women than men. Data from registry analyses and clinical trials [8] suggest that clinicians should abandon the 'silo thinking' when it comes to diagnosing VTE by considering integration of cardiovascular risk factors. These common risk factors include: obesity, smoking, hypertension, Dyslipidemia, diabetes mellitus, and metabolic syndrome. Obesity is associated with doubling in the VTE risk. Other studies showed that obesity in women triples the risk of unprovoked PE. Diabetes has been associated with 42% of increased risk of DVT or PE. The Wells score [9, 11] and the revised

Geneva score [11] models are only used during physical examination after exhibition of symptoms in the patient. The approach used in [5] to solve the semantic gap problem is a concept-based approach that uses medical domain knowledge from the SNOMED-CT ontology while for concept extraction MetaMap is utilized. Most recently initiative towards making formal ontology of cardiovascular disease has been taken.

### 3 Semantic Framework to Predict VTE

Our framework consists of four major phases: Feature extraction, feature classification, risk factors scoring, and calculation of the clinical probability, as showing in Figure 1.

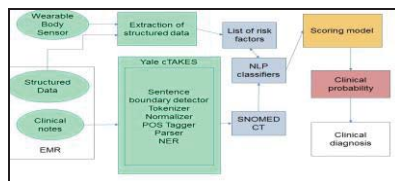


Fig. 1. Semantic model for predicting first occurrence of VTE

#### 3.1 Feature Extraction

Clinical natural-language-processing systems annotate the syntax and semantics of clinical text. YTEX simplifies feature extraction. Our model extracts data from two major sources: electronic medical record and wearable body sensor(s). The novel approach in the extraction of unstructured data is to semantically analyze the free form text and find the medical annotations concerning the cardiovascular disease risk factors.

#### 3.2 Feature Classification

In the classification process, we consider the use of the medical ontology library SNOMED CT to identify and refine the extracted list of terms as causes of VTE. For example, a doctor might write in the clinical note an abbreviation of a clinical term such as T2 or a medical term mellitus or a simple expression high blood glucose, which all can be classified under one risk factor listed as “diabetes”.

#### 3.3 Scoring Model and Clinical Probability

In table 1 we list the risk factors for symptomless unprovoked VTE and their relative scores. For the clinical probability: Score < 3 then probability < 30% low risk  
 3 < Score < 5 then 30% < probability < 50%: precaution  
 Score > 5: immediate admission for venography and initiation of blood thinner such as heparin to dissolve any formed clots.

### 4 Conclusions

Our model contributes to the field of preventive medicine by predicting the occurrence of first time symptomless unprovoked VTE using a new semantic approach that extracts risk factors features from structured data and unstructured text

of EMRs. More work and analysis will be completed in the coming months to publish the results of our study. As future work, we plan as well to broaden our framework and apply it to other diseases. Our work will be completed in collaboration with Henry Ford Hospital in West Bloomfield, Michigan.

Table 1. Risk factor scores by data sources

EMR structured	EMR unstructured	WBS
Age < 45 +2, >45 +1	Inherited gene +3	Obesity or BMI +2 (calculated)
Weight	Vasculose veins +2	Heart rate
Height	Air pollution +1	Blood pressure
Chronic renal disease +1	Pregnancy +1	Sleep hours +1
Inflammatory bowel disease +1	Postpartum +1	
Hypertension +1	Contraceptive +2	
Dyslipidemia (metabolic syndrome) +1	Long travels +2	
Diabetes +1	Stress +2	
Smoking +1	Beer +1	
Chronic heart disease +1		
Neurological disease +1		
Chronic lung disease +1		
Liver disease -1		

### 5 References

- [1] X Zhou, H Han, I Chankai, A Prestrud, A Brooks “Approaches to text mining for clinical medical records”. In: Proceedings of the 2006 ACM symposium on Applied computing: ACM; 2006. p. 235-239.
- [2] K Tomaneck, J Wermter, U Hahn. “A reappraisal of sentence and token splitting for life sciences documents”. Studies in health technology and informatics 2007;129: 524.
- [3] S Ananiadou, J McNaught. “Text mining for biology and biomedicine”. Artech House Boston, London; 2006.
- [4] HD Tolentino, MD Matters, W Walop, B Law, W Tong, F Liu, P Fontelo, K Kohl, DC Payne. “A UMLS-based spell checker for natural language processing in vaccine safety”. BMC medical informatics and decision making 2007;7: 3.
- [5] B Koopman, P Bruza, L Sitbon, and M Lawley. “Towards semantic search and inference in electronic medical records: An approach using concept-based information retrieval”. Australas Med J. 2012; 5(9): 482-488. Published online Sep 30, 2012. doi: 10.4066/AMJ.2012.1362 PMID: PMC3477777
- [6] A. Elias, L. Mallard, M. Elias, C. Alquier, F. Guidolin, B. Gauthier, A. Viard, P. Mahouin, A. Vinel, and H. Boccalon, “A single complete ultrasound investigation of the venous network for the diagnostic management of patients with a clinically suspected first episode of deep venous thrombosis of the lower limbs,” Management, vol. 1, pp. 5-10, 2003.
- [7] L. A. Castellucci, C. Cameron, G. Le Gal, M. A. Rodger, D. Coyle, P. S. Wells, T. Clifford, E. Gandara, G. Wells, M. Carrier, and others, “Efficacy and safety outcomes of oral anticoagulants and antiplatelet drugs in the secondary prevention of venous thromboembolism: systematic review and network meta-analysis,” BMJ, vol. 347, 2013.
- [8] G. Piazza and S. Z. Goldhaber, “Venous Thromboembolism and Atherothrombosis An Integrated Approach,” Circulation, vol. 121, no. 19, pp. 2146-2150, May 2010.
- [9] S. Z. Goldhaber and H. Bounameaux, “Pulmonary embolism and deep vein thrombosis,” The Lancet, vol. 379, no. 9828, pp. 1835-1846, May 2012.
- [10] J. H. Scurr, S. J. Machin, S. Bailey-King, I. J. Mackie, S. McDonald, and P. D. C. Smith, “Frequency and prevention of symptomless deep-vein thrombosis in long-haul flights: a randomised trial,” The Lancet, vol. 357, no. 9267, pp. 1485-1489, May 2001.
- [11] P. S. Wells, D. R. Anderson, M. Rodger, M. Forgie, C. Kearon, J. Dreyer, G. Kovacs, M. Mitchell, B. Lewandowski, and M. J. Kovacs, “Evaluation of D-Dimer in the Diagnosis of Suspected Deep-Vein Thrombosis,” New England Journal of Medicine, vol. 349, no. 13, pp. 1227-1235, Sep. 2003.
- [12] JA Heit, MD Silverstein, DN Mohr, TM Petterson, W O’Fallon, L Melton “Risk factors for deep vein thrombosis and pulmonary embolism: A population-based case-control study,” Arch Intern Med, vol. 160, no. 6, pp. 809-815, Mar. 2000.
- [13] S. Eichinger, G. Heinze, L. M. Jandeck, and P. A. Kyrle, “Risk Assessment of Recurrence in Patients With Unprovoked Deep Vein Thrombosis or Pulmonary Embolism The Vienna Prediction Model,” Circulation, vol. 121, no. 14, pp. 1630-1636, Apr. 2010.



# A Metaverse Client of an Assisted Living Support Service System for Extended Family

Mi-Kyong Han <sup>1</sup>, Jong-Hyun Jang <sup>2</sup>

<sup>1,2</sup> Realistic & Emotional Convergence Service Research Lab,  
Electronics and Telecommunications Research Institute,  
Daejeon, Republic of KOREA

**Abstract** - *In the Internet of Things (IoT), devices gather and share information directly with each other and the cloud, making it possible to collect record and analyze new data streams faster and more accurately. These Advancements in embedded information and communication technologies present enormous potential for the intensified healthcare support of senior citizens at home. By employing these technologies in the home, senior citizens are able to live independently for a longer period of time, helping to reduce costs and the need for additional caregiver resources in the process. These techniques are applied to healthcare or safety home area on a commercial scale. But, there are many potential problems that the analysis of home situation with sensed data is not covered with the various difference homes living pattern of each person. And that's why it's extremely important to determine precisely what it has happen.*

*To solve this problem, we will address the assisted living support service system (defined as a MACH: Metaverse based Assisted living support system for the Connected Home) based on a metaverse client dedicated to the smart home. MACH system supports the different living pattern of different home. And, it is able to access to the monitoring information with metaverse client. It can be support the extended family members convenient to use with a client interface.*

**Keywords:** Assisted Living Service, Metaverse Client, Healthcare System, Remote Monitoring, Living Pattern Analysis, Context Awareness

## 1 Introduction

The Internet of Things (IoT) is the logical further development of today's internet. Technological advancements lead to smart objects being capable of identifying, locating, sensing and connecting and thus leading to new forms of communication between people and things and things themselves. Advancements in the Internet of Things (IOT) present enormous potential for the intensified smart home. Smart home concept may provide a variety of services dealing with multimedia (streaming video and sound, reading emails), comfort (ambient lighting, heating) as well as health supervision (blood pressure monitoring, heart attack prevention). One of the smart home services is assisted living

service that it is a residential alternative to nursing home care, support services and health care, as needed. Assisted living service is designed for individuals who require assistance with everyday activities such as meals, medication management or assistance.

Although their objectives are some different, these approaches are very similar technically speaking: the house continuously monitors the environment in order to provide better-adapted services to the inhabitant. Assisted living service has become more important owing to the increasing number of a solitary elderly or a single household. Assisted living service can support the analysis of home situation with sensed data is not covered with the various difference homes living pattern of each person. And the system has to determine precisely what it has happen to person.

Some of the existing context aware medical applications are: In VOCERA COMMUNICATION SYSTEM [1] experimentation was done in St. Vincent Hospital, Birmingham, USA. STROKE ANGEL [2] experiment was conducted at Bad Neustadt, Germany. (November 2005-May 2008). One more study was conducted for a period of 8 months in a mid-size public hospitals in the city of Ensenada, Mexico [3]. A decision-level data fusion technique for monitoring and reporting critical health conditions of hypertensive patients [4].

In this paper, we propose an assisted living support service system (defined as a MACH:Metaverse based Assisted living support system for the Connected Home) . MACH system supports the different living pattern of different home. And it supports a convenient to use access interface based on metaverse client. Metaverse client is able to compose of extended home from many nuclear home and it can be accessed by extended family members.

## 2 MACH SERVICE SYSTEM

In this section, we suggest a MACH service system that supports an assisted living service using the metaverse client for extended family. Figure 1 shows architecture of a suggested MACH system, which is aware of changes user's living pattern situation in smart home environments. As shown in Figure1, a suggested system is composed of several components interacting towards the others.

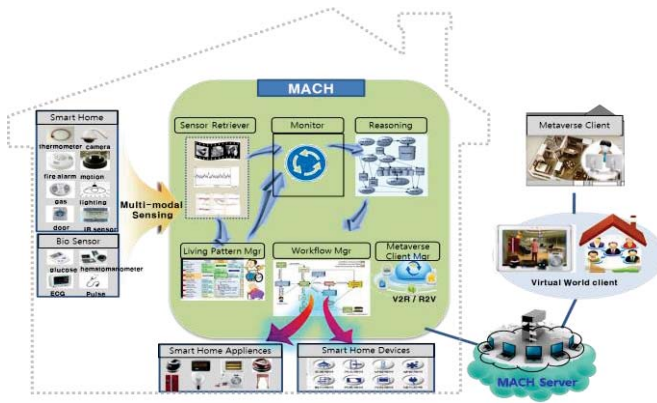


Figure I. The components of proposed system for an assisted living support service system

- **Sensor Retriever:** requests and receives sensed information from sensors of smart home. There are many kind of sensors in reality can be listed as light, temperature, humidity, camera, location, person, etc.
- **Monitor:** monitoring the user's status referencing the sensed status information and user's living pattern information.
- **Reasoning:** queries and concluded the appropriate actions to the current context.
- **Living Pattern Manager:** updates and keep up the user's living pattern with the latest information for monitor.
- **Workflow Manager:** keep up the workflow information for appropriate control on current context with the various difference homes living pattern of each person.
- **Metaverse Client Manager:** controls the virtual home environments such as extended family member, home appliances and configuration information, etc.

Those components have relationships to the others. Monitor and reasoning part receives data from sensor retriever as well as creates the event and sends to workflow manager. The workflow manager controls the home appliances or sends the status information to specified extended family members to check the user's status referencing the pre-defined workflow by users.



Figure II. The interface components of MACH client

The suggested system uses the metaverse client to control the home appliances and check the user's living pattern and status by home owner as well as the other extended family members. Figure 2 shows the interface components of MACH client. MACH client can be specified with the virtually extended family home environment, which is actually composed of many of nuclear family home units. Users establish the virtual home environment and control the virtual home status using the various client interfaces (R2V:Real to Virtual, V2R:Virtual to Real, V2V:Virtual to Virtual) for personal privacy and simplicity of the client interface.

### 3 Conclusions

In this paper, we propose an assisted living service system MACH to support user's healthcare by adopting personal living pattern and metaverse client. Through this system, it can be provides the assisted living care as well as the access interface for extended family members to protect privacy each other with the virtual client interface.

In the further works, we aim to test our system using the various test scenarios to get the vision up and verify our system. We will define the various test scenarios and construct the experiment environment. Through an experiment with the selected test scenario, we will show that the system can be served the assisted living support service or home security service to provide the easy to access interface the senior citizen and others. And we will running our next steps in order to more accurately detect specific resident situations through the research in the area of behavior patterns.

### 4 References

- [1] Vince Stanford, "Beam me up, Dr. McCoy". IEEE Pervasive Computing Magazine 2(3), pp. 13–18 (2003).
- [2] Carsten Orwat, Asarnusch Rashid, Carsten Holtmann, Michaela Wolk, Mandy Scheermesser, Hannah Kosow, Andreas Graefe, "Adopting Pervasive Computing for Routine Use in Healthcare". IEEE Pervasive Computing, Vol. 9 No. 2, pp.64-71,(April-June 2010).
- [3] Jesus Favela, Monica Tentori, Luis, A.C., Victor M.G., Elisa B.M., Ana I.M.G, "Estimating Hospital Work Activities in Context-Aware Healthcare Applications". Mexico City Pervasive Health Conference and Workshops, pp.1-10 (2006)
- [4] Alessandro Copetti, Orlando Loques, Leite, J.C. B., Thais P. C.Barbosa, da Nobrega, Antonio C.L. , "Intelligent Context-Aware Monitoring of Hypertensive Patients". Pervasive Computing Technologies for healthcare, pp.1-6, 2009. Pervasive Health 2009, pp. 1-3 (April 2009).

### Acknowledgment

This research was supported by the IT R&D programs of the ministry of science, ICT and Future Planning (MSIP) [10043430, Development of Metaverse based Collaborative Family Safety Service Virtualization Technology for Remotely Resizer Family Members].

# Qualitative Analysis of the Acceptance of EMR Systems

ABSTRACT/POSTER PAPER

Raghd El-Yafouri<sup>1</sup> and Leslie Klieb<sup>1,2</sup>

<sup>1</sup>Grenoble Ecole de Management, Grenoble, France

<sup>2</sup>Webster University Thailand, Bangkok Campus, Bangkok, Thailand

**Abstract** - This qualitative case study discusses acceptance and problems of physicians with electronic medical record systems. The results of this case study agree with 196 mainly negative comments left by respondents to a quantitative survey on the same topic among small practices in the USA.

**Keywords:** Electronic Medical Record systems, Acceptance among physicians, Qualitative Case study.

## 1 Introduction

El-Yafouri and Klieb [1,2] conducted research on diffusion and barriers of electronic medical records (EMR) in the USA in small size medical practices. The quantitative part was done with a survey [2]. Part of this research was also a qualitative multiple case study. We present for the first time some preliminary results of this qualitative study here. Additionally we discuss remarks made as reply to an open question on this survey. Results are preliminary and we plan to continue research.

## 2 Methodology

The qualitative results of this multiple case study were obtained by semi-structured interviews. The sampling strategy was a purposeful sampling strategy [3] (page 125, 75). Participants were selected to obtain maximum variation [4]. A fair amount of saturation was accomplished. In our analysis we followed [5] Wolcott (1994) in that we identified patterns and contextualize this with the existing literature and quantitative findings (compare [3] page 149).

## 3 Informants

Our informants were: 1) Dr. A., MD, Ph.D., full professor of clinical medicine, treating heart failure, heart disease from diabetes and delivering other chronic long-term care, 31 years' experience, 2) Dr. B. and Dr. C., physicians with 6 and 4 years' experience in an ICU and 3) Dr. D., gynecologist in a private practice, with 12 years of experience. (For privacy reasons not the actual names and affiliations are given here). Results are collected in table 1.

## 4 Conclusions

Preliminary conclusions for a grounded theory approach are: Acceptance is larger when doctors have never worked with a manual system and cannot compare the system, do not have long-term care patients and therefore do not have to rely on the entered data at later visits, are younger and more tech-savvy, and use a system that is more advanced in its data-sharing aspects. Doctors who have to provide more care to long-term care patients or rely on many small details, like in cardiology, and who had higher expectations or require that the system is doing better than a manual system are not accepting of electronic systems. Doctors struggle with adjusting the granularity of their notes to the systems. None of the doctors brought up the possibility of doing more research with the collected big-data.

In the survey 196 additional comments were made, showing how important the issues are for MDs. Less than 10% were from physicians who liked their systems. All aspects are already covered above, therefore saturation was achieved, except from one comment that vendors promise too much. This research will be continued and expanded.

## 5 References

- [1] Raghd El-Yafouri, Electronic Medical Records Adoption and Use, Understanding the Barriers and the Levels of Adoption for Physicians in the USA. Dissertation Grenoble Ecole de Management, March 2015.
- [2] Raghd El-Yafouri and Leslie Klieb, Electronic Medical Records Adoption and Use: Understanding the Barriers and the Levels of Adoption for Physicians in the USA. 2014 Healthcom IEEE conference, Natal, Brazil.
- [3] John W. Creswell, Qualitative Inquiry and Design, SAGE, 2<sup>nd</sup> ed., 2007.
- [4] Matthew A. Miles & Michael B. Huberman, Sourcebook of Qualitative Data Analysis, SAGE, 1994.
- [5] Harry F Wolcott, Transforming qualitative data: Description, Analysis, and Interpretation. SAGE 1994.

Subject	Category	A	B/C	D
Using EMR system	Attitude	Dislike	Like	Dislike, partner dislikes too, diligent
Data sharing between systems and departments of the office/organization	EMR Level	No	Yes	Not compatible with scheduling system
Data sharing between different medical institutions	EMR Level		Beneficial	In theory easier, Scary, Too much transparency, Labs not connected
Paperless environment	EMR Level	No	No	No; Insurance companies require paper copies
Alerting capabilities	EMR Level	Only when logged in		
Electronic prescription	EMR Level	Not implemented		Office and pharmacy systems are incompatible
Accessibility of electronic records	Technical	Should be anytime		Less likely to lose data with EMR Impacted when network down
Error reduction	Technical	No	Yes	Yes
Computer knowledge	Technical	Required to use EMR		Required to use EMR
Ease of use	Technical	Difficult	Easy	Cumbersome
Note taking and data entry	Technical	Can't jot notes Easier with dictating and scanning		Requires more time. With multiple systems inconsistencies. Cannot be done real time. Paper lab results and other papers require scanning
Training and user support	Technical	Not sufficient		Not sufficient
Long term benefits of EMR	Psychological	Not aware		Not aware
Impact of EMR adoption	Psychological	Older MDs will retire		Older MDs will retire instead of using
Input medical professionals in system design & selection	Knowledge	No input		No input
Saving time	Time	Requires more time	Saves time	Requires more time
Staffing needs	Organizational			
Role of champion	Organizational	Does not help	Did not have a champion	Is important
Workflow in my office/organization	Organizational		Enhanced	Easier to re-engineer in a start-up than in an established one
Transitioning and/or implementing EMR	Organizational			
Physician's personal satisfaction	Organizational	Eliminated with EMR		
Colleagues' attitude	Social/Legal	Dislike	Like	Partner dislikes too
Federal government's mandate	Social/Legal	Main reason for adoption	Main reason for adoption	Main reason for adoption
Privacy of patient records	Social/Legal	Prevents data sharing	Reduced	

# Socio-Spatial Analysis of Florida Suicide Mortality

## Poster Paper

Asal Johnson<sup>1</sup>, Joseph M. Woodside<sup>2</sup>, and Jacqueline Pollack<sup>1</sup>

<sup>1</sup>Department of Integrative Health Sciences, Stetson University, DeLand, FL, USA

<sup>2</sup>Department of Decision and Information Sciences, Stetson University, DeLand, FL, USA

**Abstract** - *In the United States suicide was the 10th leading cause of death accounting for more than 40,000 deaths in 2010 [1]. In Florida suicide was the 9th leading cause of death in 2013 where the number of suicides from 2011-2013 exceeded 8000 deaths [2]. The purpose of this study is to determine social and spatial patterns of suicide deaths in the state of Florida through cluster analysis. The project merged the aggregated number of suicide deaths files from the Florida Department of Health, census data and rural urban continuum codes to obtain neighborhood level constructs contributing to the existence of possible suicide clusters. The results of this proposed study will guide interventions that have the potential to build more cohesive, healthy and supportive networks for residents of areas that encounter higher rates of suicide.*

**Keywords:** Social, Spatial, Clusters, Suicide, Florida

## 1 Introduction and Overview

Suicide was the tenth leading cause of death in the United States as it accounted for 41,149 deaths in 2010. This is equivalent to one individual dying every 12.8 minutes from suicide in the United States [1]. In Florida, the suicide rate among males steadily increased from the age of 15 through 64 and the suicide rate among females increased from the age of 15 through 54 during 2012 [3]. Suicide was the ninth leading cause of death with more than 8,000 deaths in 2013 compared to the previous year of approximately 3,000 suicides in 2012 for Floridians [3, 4]. The increased rate of suicides warrants further investigation as the presence of cluster suicides have increased as well [4]. It has been suggested that suicide clusters might be explained through contagion, two or more suicides taking place due to the influence of a previous suicide within a close proximity in time and space [5, 6].

Previous studies and cluster analyses have tried to explain the increased prevalence of suicides by identifying contact to health care, depression and other mental illnesses, divorce rate, substance abuse, population density, socioeconomic and living circumstances in the past as risk factors [7-10]. A majority of the current literature states that rates of suicide are often higher in rural areas than in urban areas. For example, at the state level, age-adjusted male suicide rates vary from 11 deaths per 100,000 in urban areas such as Washington D.C to approximately 36 deaths per 100,000 in more rural states like Alaska and Wyoming. In Canada, as community size

decreases, the suicide rate increases. In Australia, rates of rural communities' suicide doubled and rates for youths in rural municipalities increased five-fold. Higher rates of rural suicides have also been seen in Estonia, Germany, Lithuania, Finland and numerous other developed countries throughout Europe [11, 12].

In Florida, rural is defined as an area with a population density of less than 100 individuals per square mile [3]. According to the 2012 Florida Injury Facts Surveillance data, rural communities have fewer hospitalizations from suicide or non-fatal self-harm compared to non-rural communities with age-adjusted rates/100,000. The research on suicide prevalence in rural and urban communities is inconclusive as in developing countries, like Columbia, suicide rates are higher in urban area [13]. The statistics given by the Florida Health Department also contradicts the expected higher rural suicide rates. Regardless of rural versus urban populations, suicide hotspots are common in areas with tall buildings, cliffs, railway tracts, bridges and secluded locations [14, 15, 16]. Living in a neighborhood characterized by a poor quality built environment, whether urban or rural, is associated with a greater likelihood of depression and depending on the severity, could lead to suicide [17].

## 2 Spatial Analysis

In recognition of the usual epidemiological definition of cluster, this study adopts the formal definition of cluster which refers to the patterns of location of disease cases, relative to the pattern of non-cases. In principle, since the cases are more clumped than non-cases, the difference between the two patterns is statistically recognizable. Cluster analysis has been frequently used to identify occurrence of unusual localized trends in disease patterns. This research utilizes a 'scan' statistic, which has been identified as the most powerful method for detecting local clusters. The method is based on the concept of 'windows' which are defined to contain a fixed population, and are centered on each area centroid. The algorithm identifies a significant excess of case within a predefined moving window that exhaustively searches all existing space-time locations and keeps increasing size in space time until it reaches a maximum limit. The maximum number of cases across the windows is used as a test statistic. The test statistic is based on maximum likelihood ratio statistic across all circles.

$$L = \max_j \left( \frac{Y_j}{E_j} \right)^{Y_j} \left( \frac{Y_+ - Y_j}{Y_+ - E_j} \right)^{Y_+ - Y_j} I(Y_j > E_j)$$

Here,  $Y_j$  and  $E_j$  represent the observed and expected number of cases within the window  $j$ . The indicator function  $I(.)$  becomes 1 when observed number exceeds the expected number of cases within the window, otherwise the value is 0. When the window with greatest exceedance is encountered, the sampling distribution of likelihood ratio is determined using a Monte Carlo test of cases across windows under random distribution assumption. Thus under the repeated permutation of likelihood the distribution of likelihood static the null hypothesis is developed. The result is significant at 0.05 levels if the likelihood ratio is among the top 5% of all the values. It is also possible to determine secondary clusters with lower significance level [18, 19].

### 3 Results and Future Directions

The results of the study are utilized to identify suicide clusters in Florida and compare social and spatial patterns of these clusters with low rate suicide neighborhoods. Cluster analysis was conducted through formalized statistical and geographic information systems with geographic results layered and triangulated. Several spatial areas were detected which deviated from randomness and contained a statistically significant cluster of excess suicide cases.

For future directions, epidemiological considerations and implications will be investigated including social, economic and environmental contexts including covariates for adjusted analyses. Initial assessments of clusters include review of cases, boundaries of space and time, estimated number of cases, estimate of standardized mortality ratios, statistical significance, and public communication. The dimension of social environment will be conceptualized by social integration and racial composition. Economic environment variables will be analyzed and measured by a composite variable utilizing factor analysis. The built environment will be captured by three variables: the percentage of housing structures with more than 5 units, population density, and automobile dependency. The results of this study will guide interventions that have the potential to build more cohesive, healthy and supportive networks for residents of areas that encounter higher suicide rates.

### 4 References

- [1] American Foundation for Suicide Prevention, *Suicide Death*, Available at <https://www.afsp.org/understanding-suicide/facts-and-figures>, Accessed February 2, 2015
- [2] Florida Department of Health, *Injury & Violence*, Available at <http://www.floridacharts.com/charts/InjuryAndViolence/> Accessed February 2, 2015
- [3] The FACTS/Statistics. (n.d.). Retrieved May 12, 2015, from [http://www.floridasuicideprevention.org/the\\_facts.htm](http://www.floridasuicideprevention.org/the_facts.htm)

- [4] Prevention. (n.d.). Retrieved May 14, 2015, from <http://www.floridahealth.gov/programs-and-services/prevention/suicide-prevention/index.html>
- [5] Cheung, Y. T. D., Spittal, M. J., Williamson, M. K., Tung, S. J., & Pirkis, J. (2014). Predictors of suicides occurring within suicide clusters in Australia, 2004–2008. *Social Science & Medicine*, 118, 135-142
- [6] Rezaeian, M. (2012). Suicide clusters: introducing a novel type of categorization. *Violence and victims*, 27(1), 125-132.
- [7] Cheung, Yee Tak Derek, et al. "Spatial analysis of suicide mortality in Australia: Investigation of metropolitan-rural-remote differentials of suicide risk across states/territories." *Social Science & Medicine* 75.8 (2012): 1460-1468.
- [8] Rehkopf, D. H., & Buka, S. L. (2006). The association between suicide and the socio-economic characteristics of geographical areas: a systematic review. *Psychological medicine*, 36(02), 145-157.
- [9] Sinyor, M., Schaffer, A., & Streiner, D. L. (2014). Characterizing suicide in Toronto: an observational study and cluster analysis. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, 59(1), 26.
- [10] Taylor, R., Page, A., Morrell, S., Harrison, J., & Carter, G. (2005). Social and psychiatric influences on urban-rural differentials in Australian suicide. *Suicide and Life-Threatening Behavior*, 35, 277–290.
- [11] Hirsch, J. K. (2006). A review of the literature on rural suicide. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 27(4), 189-199.
- [12] National Center for Injury Prevention and Control (NCIPC), Centers for Disease Control and Prevention. (2012b). Webbased injury statistics query and reporting system (WISQARS). Retrieved from <http://www.cdc.gov/ncipc/wisqars>
- [13] García Valencia, J., Montoya Montoya, G. J., López Jaramillo, C. A., López Tobón, M. C., Montoya Guerra, P., Arango Viana, J. C., & Palacio Acosta, C. A. (2011). Characteristics of Suicides in Rural and Urban Areas in Antioquia, Colombia. *Revista colombiana de psiquiatria*, 40(2), 199-214.
- [14] Beautrais, A. L., Gibb, S. J., Fergusson, D. M., Horwood, L. J., & Larkin, G. L. (2009). Removing bridge barriers stimulates suicides: an unfortunate natural experiment. *Australasian Psychiatry*, 43(6), 495-497.
- [15] Cox, G. R., Owens, C., Robinson, J., Nicholas, A., Lockley, A., Williamson, M., & Pirkis, J. (2013). Interventions to reduce suicides at suicide hotspots: a systematic review. *BMC public health*, 13(1), 214.
- [16] Pirkis, J., Spittal, M. J., Cox, G., Robinson, J., Cheung, Y. T. D., & Studdert, D. (2013). The effectiveness of structural interventions at suicide hotspots: a meta-analysis. *International journal of epidemiology*, dyt021.
- [17] Galea, S., Ahern, J., Rudenstine, S., Wallace, Z., & Vlahov, D. (2005). Urban built environment and depression: a multilevel analysis. *Journal of Epidemiology and Community Health*, 59(10), 822-827.-Reference number 9 (need to go back and shift all the references in the review and reference sheet down a number)
- [18] Woodside, Joseph M, Sikder, Iftikhar U. GIS Application of Healthcare Data for Advancing Epidemiological Studies. *New Technologies for Advancing Healthcare and Clinical Practice*. 2011.
- [19] Woodside, Joseph M, Sikder, Iftikhar U. Space-Time Cluster Analysis: Application of Healthcare Service Data in Epidemiological Studies. *International Journal of Healthcare Information Systems and Informatics*. 2009.

# Development of Medical 3D Printing Software for Porous Scaffold Generation

Sangbeom Lee and Eunchang Choi

Electronics and Telecommunications Research Institute (ETRI)  
1, Techno sunhwan-ro 10-gil, Dalseong-gun, Daegu, Republic of Korea  
E-mail: {sblee230, ecchoi}@etri.re.kr

**Abstract** - Recently, 3D printing technology has played an important role in a variety of medical applications. Although various types of 3D printing devices have been released, medical 3D printing software for porous scaffold structure construction has not been well-developed yet. Therefore, in this paper, we present a 3D printing software which enables porous scaffold generation. First, a 3D model which is composed of numerous triangles is discretely sliced in the z-axis from the bottom to top. Then, the intersection polygon between each slice and 3D model is obtained by searching and connecting intersection lines between the z-plane and their corresponding triangles. From the intersection polygon, a toolpath is generated considering porous scaffolds.

**Keywords:** Medical 3D printing, scaffold generation, STL slicing

## 1 Introduction

Additive manufacturing (AM), commonly called as Rapid Prototyping (RP), has been widely used in overall industries such as building, automobile, entertainment, etc. Nowadays 3D printing is now in the spotlight due to time reduction and flexibility improvement of product manufacturing [1][2]. Furthermore, it can now be applied in medical application since it makes possible customization of artificial bone and scaffold structure.

Basically, 3D printing technology handles stereo lithography (STL) file format to represent 3D geometry information. For additive manufacturing, slicing and toolpath generation is essential process. Slicing process converts 3D data to a set of planes. Similarly, toolpath generation changes 2D data to a set of lines. In order to construct porous scaffold structure for medical use, crisscross toolpath should be generated. In this paper, we create a porous scaffold generation solution with the aid of a 3D printer.

## 2 3D Model Slicing Algorithm

### 2.1 Intersection Point Detection

As a first step in 3D printing software, the 3D model is discretely sliced in z-axis from the bottom to top after STL file is loaded and forms triangular mesh structure. For each z-plane, triangles are evaluated whether they intersect the z-plane or not. Unlike some researches classify the intersection into many cases, we only consider two cases that the triangle has no intersection or two intersection points for simplicity. Fig. 1 shows two intersection cases between z-plane and triangle.

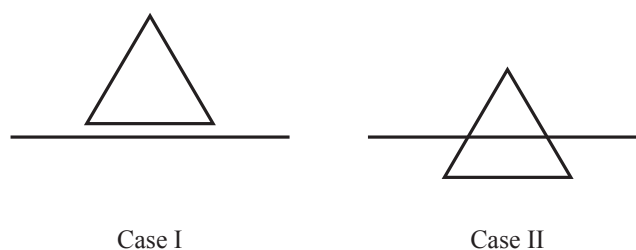


Fig. 1. Intersection cases between z-plane and triangle

After we find two intersection points and form a line segment for entire triangle, we can obtain the intersection polygons which are composed of intersection line segments for each z-plane. Fig. 2(b) shows the slicing result for the 3D model of lower jaw bone. Notice that the number of slice is dependent on the slice thickness.

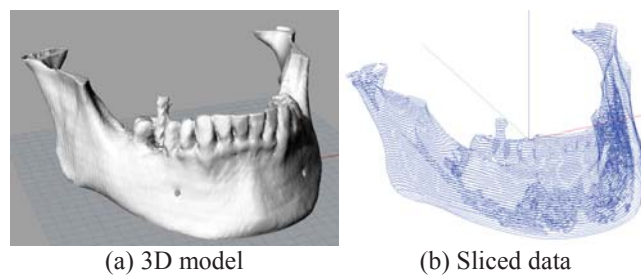


Fig. 2. Slicing result for 3D model of lower jaw bone

### 3 Toolpath Generation for Porous Scaffold

In order to fill the interior of the intersection polygons, toolpath should be generated. The general porous scaffold design forms crisscross toolpath that has a space between lines. Fig. 3 shows the general scaffold structure.

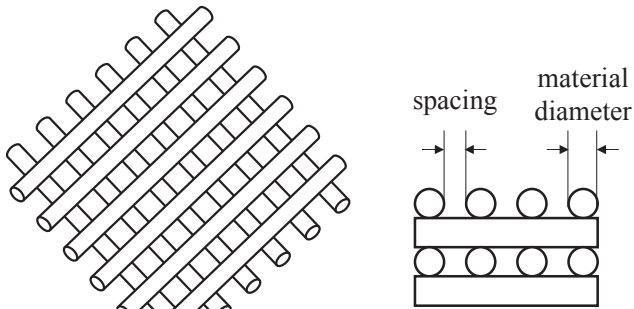


Fig. 3. Medical scaffold structure

Therefore, the toolpath is generated following the scaffold structure. First, we find the intersection points between intersection polygons and x-axis (or y-axis) parallel lines. Here, the space between lines is dependent on the pore size and the material diameter. Then, we determine toolpaths so that they are inside the polygon. This process is repeated from bottom to top slices. Notice that the x-axis and y-axis parallel lines are used by turns. Finally, we obtain toolpaths for the entire slice.

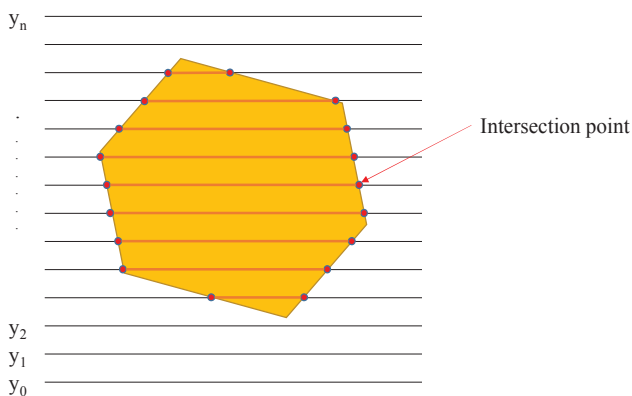


Fig. 4. Toolpath generation process

### 4 References

[1] M. Eragubi, "Slicing 3D CAD Model in STL Format and Laser Path Generation," *International Journal of Innovation, Management and Technology*, vol. 4, no. 4, pp. 410–413, Aug. 2013.

[2] R. C. Luo and J. H. Tzou, "Implementation of a New Adaptive Slicing Algorithm for the Rapid Prototyping Manufacturing System," *IEEE/ASME Transactions on Mechatronics*, vol. 9, no. 3, pp. 593–600, Sep. 2004.



## **SESSION**

# **LATE BREAKING PAPERS: HEALTH INFORMATICS AND MEDICAL SYSTEMS**

**Chair(s)**

**TBA**



## Automatic Cephalometric Landmark Detection: A Literature Review

José de Jesús Montúfar<sup>1</sup>, Marcelo Romero<sup>1</sup>, Brissa Jiménez<sup>2</sup> and Vianney Muñoz<sup>1</sup>

<sup>1</sup>Department of Engineering and <sup>2</sup>Department of Odontology

Autonomous University of the State of Mexico

{jdmontufart, mromeroh, bijimenezv, vmunozj}@uaemex.mx

### Abstract

*In this paper, a literature review on different techniques used for automatic cephalometric point or landmark detection is presented. Cephalometric analysis is a conventional planning process in various types of maxillofacial surgery, commonly done by manual identification on paper. Addressing the need for a more streamlined process and better treatment for patients, we believe that an automated cephalometric analysis save time and improve reproducibility. A comprehensive review of the algorithms used for automatic cephalometric landmark detection, advantages, disadvantages and clinical accuracy is required. Techniques discussed here are classified as model-based, knowledge-image based, soft-computing and hybrid techniques. This review is encouraging our final aim in implementing a computer-assisted cephalometric system for maxillofacial surgery planning.*

**Keywords:** Automatic cephalometric points detection, maxillofacial surgery planning, cephalometric landmarks.

### 1. Introduction

A cephalometric analysis is a procedure in which human skull measures are defined from lateral radiographs. A cephalometric analysis or just cephalometry is a standardized method that is part of the planning process in different types of maxillofacial surgery. This study derives records to be considered for a definitive diagnosis not only in orthodontics, but also in other areas of dentistry [20]-[24].

Three approaches are currently identified for cephalometric analysis: Manual, computer-assisted and automatic.

Manual cephalometry is a commonly use approach by practitioners, in which they draw on a superimposed acetate, cephalometric radiograph points and reference planes to get measurements between them.

The computer-aided approach is when cephalometric landmarks are manually localized in medical images (e.g. X-ray) for a computer to do a more sophisticated analysis.

In the automatic approach, a computer algorithm locates the cephalometric points and run the analysis from a digital X-ray. Then, an automatic cephalometric analysis is aim to: reduce working time, improve accuracy of cephalometric landmarks locations and reduce errors caused by subjective judgments.

In this paper we present a literature review on cephalometric landmarks detection, as a first step in a most important contribution on virtual surgery planning (VSP) in maxillofacial treatments [17].

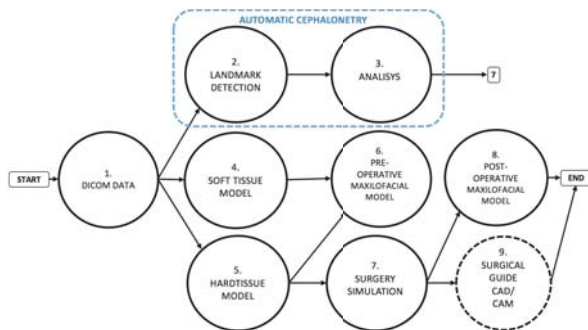
A cephalometric analysis is the first step in VSP and it is important because in here, the geometry of the maxillofacial region is defined for the rest of the surgery process. Figure 1 presents a proposed scheme for a VSP complete system, where cephalometry is an initial procedure of computerized surgical planning.

#### 1.1 Cephalometric landmarks

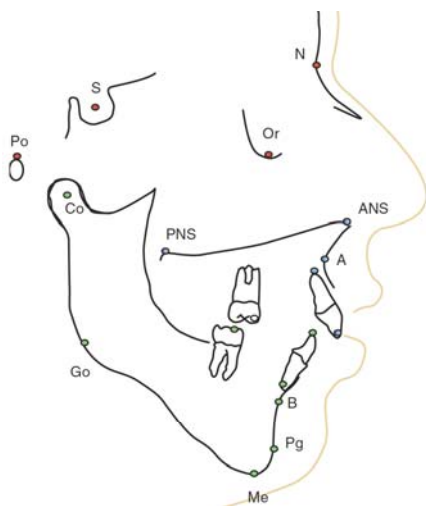
The cephalometric landmarks are defined as well recognizable points on a lateral radiograph, representing the hard tissue anatomical structures, which are used as

reference points for construction of different lines or planes to get cephalometric measurements (see Figure 2).

According to Ferraro [13], for maxillofacial surgery, there are different methods of cephalometric analysis, e.g. Rickett, Down, Tweed, Bjork, and Jarabak; and over a hundred cephalometric points. Therefore, it is necessary to choose a method and the set of points to be used.



**Figure 1.** Diagram for a maxillofacial VSP system, based on the imported soft and hard tissue patient data. Including shape and face surface color a composite model (maxillofacial model) is built and the surgical simulation is performed on 2D and 3D using software. Physical objects such as surgical guides (implant or splint) can be designed virtually and manufactured to be used inside the operation room.



**Figure 2.** Cephalometric key points [13].

Ferraro [13] has proposed cephalometric anatomical planes and a set of key landmarks: *Sella* (S), *Nasion* (N), *Orbitale* (Or) and *Porion* (Po) defining the cranial base and the Frankfurt horizontal plane, all of them are the fixed or reference structures. The *Posterior*

*Nasal Spine* (PNS), *Anterior Nasal Spine* (ANS), and point (A) define the maxilla. Point (B), *Pogion* (Pg), *Menton* (Me), *Gonion* (Go), and *Condyle* (Co) define the mandible. In addition, the points located in the incisive tips and roots and at the top of the first lower molar. These twelve cephalometric landmarks are anatomically described as follows [13]:

- a) **Sella** (S): The geometric center of the pituitary fossa (sella turcica).
- b) **Nasion** (N): The intersection of the internasal and frontonasal sutures.
- c) **Orbitale** (Or): The lowest point on the inferior orbital margin. (bilateral)
- d) **Porion** (Po): The most superior point of the outline of the external auditory meatus.
- e) **Posterior Nasal Spine** (PNS): The most posterior point on the bony hard palate.
- f) **Anterior Nasal Spine** (ANS): The tip of the bony anterior nasal spine at the inferior margin of the piriform aperture.
- g) **A Point** (A): The most posterior midline point on the curvature between the ANS and prosthion.
- h) **B Point** (B): The most posterior midline point on the bony curvature of the anterior mandible.
- i) **Pogonion** (Pg): The most anterior point on the contour of the bony chin.
- j) **Menton** (Me): The most inferior point of the mandibular symphysis.
- k) **Gonion** (Go): The most posterior inferior point on the outline of the angle of the mandible.
- l) **Condylion** (Co): The most superior posterior point on the head of the mandibular condyle.

A digital x-ray is needed for automatic detection of cephalometric points, a work began decades ago by Cohen *et al.* [14]. Nowadays, current algorithms can be categorized according to their approach [27]:

- a) Knowledge based
- b) Model based
- c) Soft-Computing
- d) Hybrids

Knowledge-based algorithms are relatively easy algorithms to implement and well

studied by the image processing research community. However, their disadvantage is a low performance with noisy and low quality images.

Model based algorithms have the advantage of being invariant to scale, rotation, translation and shape; however, a large set of training data is required and a low accuracy is expected when using low quality images.

Soft-Computing algorithms tend to be the current approach to address automatic cephalometric landmark detection when confronting variations and noise within input data. However, these algorithms also require a large training data set.

In this paper some of the relevant algorithms for automatic cephalometric landmark detection are described, indicating detected points, data sets and evaluation methods.

A lot of variability among the literature methods is found, nevertheless, for each method we are clearing advantages, disadvantages and measurements' accuracy for comparison. Also, we are assessing effectiveness in terms of the number of detected cephalometric landmarks and the number of X-rays used.

The rest of this paper is as follows. Section 2 provides background definitions. Section 3 presents a literature review on cephalometric landmarks detection. Finally, Section 4 concludes this paper.

## 2. Background

In the current approach to computer-assisted surgery (CAS) and virtual surgery planning (VSP) for maxillofacial treatments, there are two important stages: planning and simulation [17].

The planning stage is divided in two main activities: detection of cephalometric landmarks and analysis. In this stage, landmarks are detected and traced within a patient's X-ray, after that, fixed and movable planes are drawn. Literature presents, three

types of cephalometric points or landmarks [16]:

- a) Points that can be extracted directly from contours of the face, bone, orbital or ear.
- b) Points that are the center of a region, for example, *Sella*, located in the center of a *circle* on the pituitary fossa.
- c) Points that can not be extracted directly, because additional knowledge is needed, e.g. *Gn*, which is a midpoint of the curve between *Me* and *Pg*.

Computationally, detection and identification of these points is not an easy task, because it requires a specific study to address different problems that delay or hinder the analysis of a digital radiography [13], such as:

- **Position:** Position variations and errors occur when the protocol for capturing X-ray is not completely followed, resulting in rotations or inclinations that hinder the location of cephalometric points.
- **Noise:** Typically some areas in the X-rays can cause occlusion and noise by interference with brackets, crowns, and fillings.
- **Image conditions:** During the process of generating 2D cephalometric images, problems such as loss of information when scanning or in the volumetric reconstruction are found.

Each problem above involves a number of independent factors which make difficult to use common databases to evaluate cephalometric landmarks detection techniques.

## 3. Automatic Cephalometric Landmark Detection

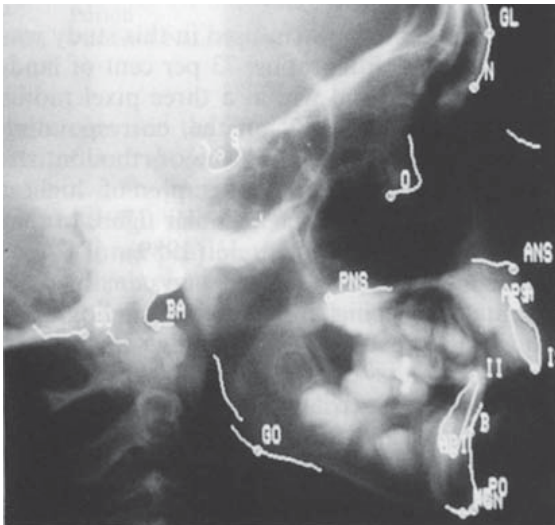
Literature presents different approaches for automatic detection of cephalometric

landmarks. We now briefly describe some of the most representative literature.

In knowledge-based algorithms, Parthasarathy *et al.* [9] detects certain cephalometric points from a pyramid of digitized X-ray. The algorithm works by searching the salient features in the images with lower resolution, which uses a priori knowledge of facial structure and then locate points of interest, moved to the higher resolution images for a more accurate location. A set of five variable quality digitized radiographs was used and pre-processed to improve contrast and edge enhancement. Accuracy for 10 landmarks was clinically acceptable in this work.

Tong *et al.* [11] uses a model of appearance characteristics and edge detection and can properly locate 17 landmarks.

Forsyth *et al.* [12] describe an automatic identification of 19 cephalometric landmarks in digital radiography and concludes that the system is less accurate than a manual study by noise in the input data, a result of an interpretation of his system is shown in Figure 3.



**Figure 3.** Example of an image interpreted by the automatic system by Forsyth *et al.* [12] for the image-knowledge automatic cephalometric landmark detection approach.

Ren *et al.* [16] proposes an *image layer* method to identify cephalometric landmarks based on multi-level knowledge. This method

builds a gray scale distribution from characteristic points of digital photographs of x-rays with two marked artificial landmarks. Then, it uses color to segment the image into bone and muscle. After that, landmarks on the bone's contour segment are directly extracted.

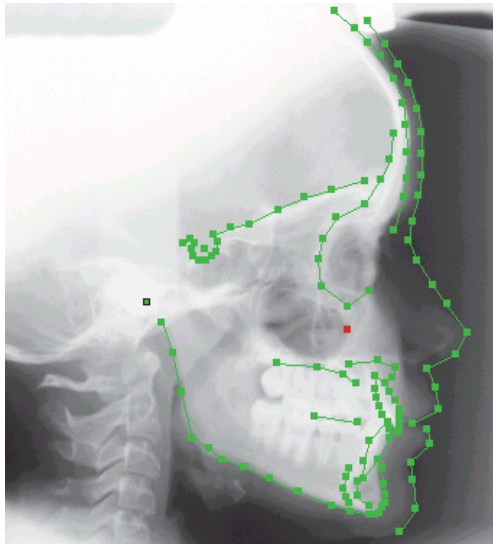
Rudolph *et al.* [2] presented a method for interpreting images with spatial spectroscopy (SS), in four steps: Eliminate noise, connect pixels of contours, tag contours and find cephalometric landmarks according to the position and relation to labeled edges. The process involves the convolution of an image with a Gaussian filter and through the first derivative images, 75 different intensity characteristic pixels are obtained and those that probabilistic are not landmark-candidates are deleted. Finally, a comparison was made between manual and automatic identification of 15 landmarks in 14 test images without significant statistical difference, concluding that the SS technique is an important step in advances for a completely automated cephalometry.

Hutton *et al.* [28] uses an active shape model (ASM), of gray level appearance and the spatial relationship of the points that define the target. This process matches the model (defined by 137 points) as a deformable template to the bone structure in X-ray for locating the landmarks of interest. The training set consists of 63 radiographs. PCA was performed to reduce high data dimensionality. It is reported that this method was not sufficiently accurate for automatic cephalometry.

Cardillo *et al.* [18] used sub-image matching on hand-selected landmarks for 40 scanned 8-bit images, and a training set of 20 landmarks. Is based on gray-scale mathematical morphology with a target recognition algorithm.

Saad *et al.* [19] uses an Active Appearance Model (124 points) and Simulated Annealing over 20 radiographs for training and 7 for testing. Results are 85% in accuracy in landmark identification. Figure 4 shows an

output of this algorithm.



**Figure 4.** Output of AAM and Simulated Annealing as an example of the use of a model for cephalometric landmark detection [19].

Chakraborty et al. [25] used support vector machine (SVM), 70 images in training set and 40 images for test. The accuracy achieved for locating 8 landmarks was 95% with 5 mm tolerance.

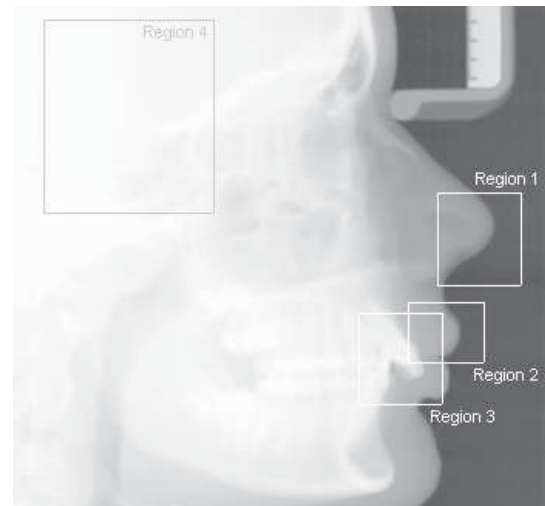
Innes et al. [5] describe a process for automatic anthropometry and cephalometry using pulse coupled Neural Network (PCNN) and the Hough transform for landmark identification. A dataset of 109 images was used and 93.7% in accuracy for identifying one landmark on soft tissue and 36.6% for two landmarks in skull are reported.

Ciesielski et al. [4] used genetic programming for four landmarks using 100 images (see Figure 5).

Vučinić et al. [10] used active appearance models (AAM), where the statistical model of the shape and greyscale converge on changes in the region it covers. It was trained with 60 digital radiographs obtaining an average accuracy of 1.68mm; they reported a 61% landmarks identified with an error lower than 2mm and a 95% landmark identification with an error lower than 5mm.

Table 1 classifies different cephalometric landmark localization techniques,

summarizing the number of landmarks, approaches and number of digital images used by each technique. Alternately, we refer the reader to Leonardi et al. [27] where other hybrid approaches, not included in this paper, are described.



**Figure 5.** A digital X-ray depicting four regions containing the mid nose, upper lip, incisal upper incisor and *Sella* cephalometric landmarks used in the genetic programming by Ciesielski et al. [4].

### 3.1. Discussion

Advances in imaging and its protocols, have increased the demand for specialists in the medical field to automatically analyze and diagnose, procedures that were previously done manually [3], [6], [8]. This has increased interest in 3D virtual surgery planning (VSP) and then in automatic cephalometric analysis, which is performed by using more than a single radiography, e.g. Moshiri et al. [35] cephalometry method for simulating lateral cephalogram with cone-beam computed tomography (CBCT), as shown in Figure 6.

There is a large variation in localization accuracy in the literature, for example some hybrid approaches are suffering to localize the *Sella*, *Porion*, and *Gonion*; unlike other approaches that found them more precisely. Hence, automatic detection of cephalometric points is still an open area for research.

**Table 1.** Automatic cephalometric landmark detection classification.

Author	Landmarks	Mean error (mm)	Method Technique	X-ray Imgs.
<b>1) Images and knowledge based</b>				
Ningrum et al. [15]	10	2.06	Projected Principal-Edge Distribution	14
Parthasarathy et al. [9]	9	1.33	Resolution Pyramid, Line Extraction	5
Tong et al. [11]	17	1.33	Resolution Pyramid, Edge, Knowledge extraction	5
Forsyth et al. [12]	19	1.50	Gray level value difference	10
Ren et al. [16]	24	--	Image layer	10
Cardillo et al. [18]	20	2.00	Pattern Matching	40
<b>2) Model Based</b>				
Rudolph et al. [2]	15	3.07	Spectroscopy, Pattern Recognition	14
Hutton et al. [28]	16	4.08	Active Shape Model	63
Romaniuk et al. [1]	1	1.20	Active contours	40
Saad et al. [19]	18	2.00	Active Appearance Model	27
Vučinić et al. [10]	9-37	1.68	Active Appearance Model	60
<b>3) SoftComputing</b>				
Chakrabarti et al. [25]	8	1.00	SVM	40
Ciesielski et al. [4]	4	2.00	Genetic Algorithms	36
El-Feghi et al. [7]	20	2.00	Fuzzy NN	600
Innes et al. [5]	3	2.00	PCNN	109
<b>4) Hybrid</b>				
Arulsevi et al. [30]	6,8	--	PCA y SVM	80
Liu et al. [31]	13	2.86	Edge Detection Model Fitting	38
Yang et al. [33]	16	1.03	Filtering, Knowledge Based edge tracing, Templates	11
Giordano et al. [26]	8	1.07	Cellular NN, Knowledge Based Extraction	26
Yue et al. [34]	12	2.00	Filtering, Edge tracking, ASM	86
Grau et al. [32]	17	1.03	Edge detection, Pattern Matching	20
Kafieh et al. [29]	16	5.00	ASM and LVQ-NN techniques	63

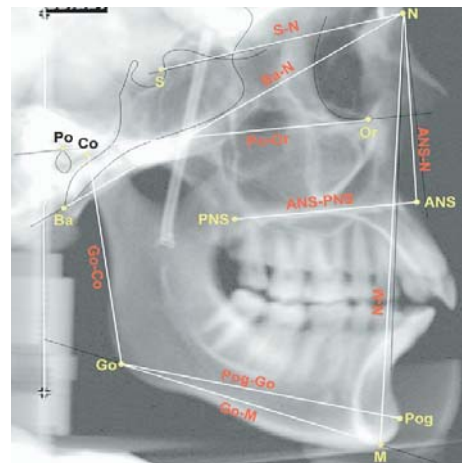
## 4. Conclusions

In this paper, a literature review on automatic cephalometric landmark detection has been presented, which is a preliminary step toward a state of the art automatic cephalometric technique.

Looking at the literature, it is complicated an objective accuracy comparison. Since, accuracy is not suitable for a total clinical diagnosis. In particular, the evolution towards an automatic cephalometry could improve

and give additional value to diagnosis, by reducing not only errors but also the planning time.

Introducing new concepts of planning maxillofacial surgery promotes the use of new technologies while leaving aside traditional methods. Virtualize surgery planning permit prior surgical intervention and prevent improvisations, while achieving a more precise procedure.



**Figure 6.** Cephalometry performed in 2D ray-sum simulated lateral cephalometric image generated from CBCT. Image from Moshiri et al. [35].

## Acknowledgements

Authors thank the Research Department of the Autonomous University of the State of Mexico and the National Council for Science and Technology (CONACYT) for their financial support, grants 3720/2014/CID and 468208, respectively.

## References

- [1] Romaniuk B, Desvignes M, Revenu M, Deshayes MJ. Shape variability and spatial relationships modeling in statistical pattern recognition. *Pattern Recogn Lett.* 2004;25:239–247.
- [2] Rudolph DJ, Sinclair PM, Coggins JM. Automatic computerized radiographic identification of cephalometric landmarks. *AmJOrthod Dentofacial Orthop.* '98;113:173–179.
- [3] Chen Y, Cheng K, Liu J. Improving cephalogram analysis through feature subimage extraction. *IEEE Eng Med Biol.* 1999;18:25–31.
- [4] Ciesielski V, Innes A, Sabu J, Mamutil J. Genetic programming for landmark detection in cephalometric radiology images. *Int J Knowl Based Intell Eng Syst.* 2003;7:164–171.



- [5] Innes A, Ciesielski V, Mamutil J, Sabu J. Landmark detection for cephalometric radiology images using pulse coupled neural networks. In: Arabnia H, Mun Y, eds. *Conf proceedings: IC-AI '02, Int Conf on Artificial Intelligence*, 24–Las Vegas, Nevada. CSREA Press; 2002; 2:511–517.
- [6] Sanei S, Sanei P, Zahabsaniesi M. Cephalograms analysis applying template matching and fuzzy logic. *Image Vision Comp*. 1999;18:39–48.
- [7] El-Feghi I, Sid-Ahmed MA, Ahmadi M. Automatic localization of craniofacial landmarks for assisted cephalometry. *Pattern Recognition*. 2004;37:609–621.
- [8] El-Feghi I, Huang S, Sid-Ahmed MA, Ahmadi M. Contrast enhancement of radiograph images based on local heterogeneity measures. Conference proceedings: ICIP '04, *International Conference on Image Processing*, 2004 Oct 24 27; Singapore. IEEE 2004;2:989–992.
- [9] Parthasarathy, Nugent S, Gregson P, Fay D. Automatic landmarking of cephalograms. *Comp Biom Res* '89;22:248–269.
- [10] Vučinić, P., Trpovski, Ž., & Šćepan, I. (2010). Automatic landmarking of cephalograms using active appearance models. *The European Journal of Orthodontics*, 32(3), 233-241.
- [11] Tong, W., Nugent, S.T., Gregson, P.H., Jensen, G. M., & Fay, D.F. '90 Landmarking of cephalograms using a microcomputer system. *Computers and Biomedical research*, 23(4), 358-379.
- [12] Forsyth, D. B., & Davis, D. N. (1996). Assessment of an automated cephalometric analysis system. *The European Journal of Orthodontics*, 18(5), 471-478.
- [13] Taub, P. J., Patel, P. K., Buchman, S. R., & Cohen, M. (2014). *Ferraro's Fundamentals of Maxillofacial Surgery*. Springer.
- [14] Cohen AM, Ip HH, Linney AD. A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis. *Br J Orthod*. 1984;11:143–154
- [15] Ningrum, I.P., Harjoko, A., & Mudjosemedi, M. (2014). Robust Cephalometric Landmark Identification on Downs Analysis. *Int J of Computer & Electrical Engineering*, 6(2).
- [16] Ren J, Lid D, Feng D, Shao J. A knowledge-based automatic cephalometric analysis method. *20th Int Conf of the IEEE Eng in Medicine and Biology Society*. IEEE '98;20(2):723–727.
- [17] Zhao, L., Patel, P. K., & Cohen, M. (2012). Application of virtual surgical planning with computer assisted design and manufacturing technology to cranio-maxillofacial surgery. *Archives of plastic surgery*, 39(4), 309-316.
- [18] Cardillo J, Sid-Ahmed MA. An image processing system for locating craniofacial landmarks. *IEEE T Med Imaging*. 1994; 13:275–289.
- [19] Saad AA, El-Bialy A, Kandil AH, Sayed AA. Automatic cephalometric analysis using active appearance model and simulated annealing. *GVIP. Int Conf on Graphics, Vision and Image Processing, Dec 19–21, Cairo, Egypt (2005)*. ICGST.
- [20] Yang J, Ling X, Lu Y, Wei M, Ding G. Cephalometric image analysis and measurement for orthognathic surgery. *Med Biol Eng Comput*. 2001;39:279–284.
- [21] Yue W, Yin D, Li C, Wang G, Xu T. Automated 2-D cephalometric analysis on X-ray images by a model-based approach. *IEEE Trans Biomed Eng*. 2006;53:1615–1623.
- [22] Stamm T, Brinkhaus HA, Ehmer U, Meier N, Bollmann F. Computer-aided automated landmarking of cephalograms. *J Orofac Orthop*. 1998;59:73–81.
- [23] Grau V, Alcaniz M, Juan MC, Monserrat C, Knoll C. Automatic localization of cephalometric landmarks. *J Biomed Inform*. 2001;34:146–156.
- [24] Liu J, Chen Y, Cheng K. Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop*. 2000;118:535–540.
- [25] Chakrabarty S, Yagi M, Shibata T, Cauwenberghs G. Robust cephalometric identification using support vector machines. Conference proceedings: *ICME, International Conference on Multimedia and Expo, 2003 Jul 6–9*. IEEE 2003;III:429–432.
- [26] Giordano D, Leonardi R, Maiorana F, Cristaldi G, Distefano M. Automatic landmarking of cephalograms by CNNs. *Lect Notes Artif Int*. 2005;3581:342–352.
- [27] Leonardi, R., Giordano, D., Maiorana, F., & Spampinato, C. (2008). Automatic cephalometric analysis: a systematic review. *The Angle Orthodontist*, 78(1), 145-151.
- [28] Hutton TJ, Cunningham S, Hammond P. An evaluation of active shape models for the automatic identification of cephalometric landmarks. *Eur J Orthod*. 2000;22:499–508.
- [29] Kafieh, R., Mehri, A., & Sadri, S. (2007, December). Automatic landmark detection in cephalometry using a modified active shape model with sub image matching. In *Machine Vision. ICMV 07* (pp. 73-78). IEEE.
- [30] Arulselvi, M., Ramalingam, V., & Palanivel, S. (2011). Cephalometric Analysis Using PCA and SVM. *International Journal of Computer Application*, 30(4).
- [31] Liu J, Chen Y, Cheng K. Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop*. 2000;118:535–540.
- [32] Grau V, Alcaniz M, Juan MC, Monserrat C, Knoll C. Automatic localization of cephalometric landmarks. *J Biomed Inform*. 2001;34:146–156.
- [33] Yang J, Ling X, Lu Y, Wei M, Ding G. Cephalometric image analysis and measurement for orthognathic surgery. *Med Biol Eng Comput*. 2001;39:279–284.
- [34] Yue W, Yin D, Li C, Wang G, Xu T. Automated 2-D cephalometric analysis on X-ray images by a model-based approach. *IEEE Trans Biomed Eng*. 2006; 53:1615–1623.
- [35] Moshiri, M., Scarfe, W.C., Hilgers, M.L., Scheetz. (2007). Accuracy of linear measurements from imaging plate and lateral cephalometric images derived from cone-beam computed tomography. *Am J of Orthodontics and Dentofacial Orthopedics*, 132(4), 550-560.

# Open Healthcare Platform using Health Information of Remote Measuring Device

Kyoungyoung So

Division of Convergence Technology Engineering, Chonbuk National University, Jeonju City, South Korea

Kwangman Ko

School of Computer and Information Engineering, Sangji University, Wonju City, South Korea

**Abstract** - *U-health that is expected to improve efficiency in medical service and monitor patients' condition with wireless communication through convergence of ICT and health care industry. A variety of health devices were rapidly spread and products were released based on many development platforms accordingly. Nevertheless, there are many development platforms that are preferred per function because of the features that each development platform has. In this paper, we suggested the development result of service and contents oriented open health care platform that comprehensively manages healthcare products created in many development platforms in a common ecosystem to build a personal health management system.*

**Keywords:** Health Care, Platform, U-health, IoT

## 1 Introduction

As u-health becomes common that monitors body condition real time in the ubiquitous environment, people are increasingly interested in promoting their health using biometric information identified by various health equipment. In particular, the concept of digital health has emerged that encompasses the followings: u-health that is expected to improve efficiency in medical service and monitor patients' condition with wireless communication through convergence of ICT and health care industry; smart health(s-health) that manages their own workout, calorie intake and sporting activities with their smart device; and, mobile health(m-health) that uses wearable and mobile devices as a means of healthcare.

The Apple's HealthKit[1] collects personal health information with external devices and applications as a platform that integrates and manages personal health information (PHI) to store and manage. The company establishes open healthcare ecosystem by attracting third party service providers in the healthcare industry. Google Fit[2], in the meantime, serves as a central storage that receives and shares health information and maintains more open platform than Apple's HealthKit. It also allows other external service providers to use the integrated PHI and access various information unlike the Apple's health application. Therefore, the other service providers can collect health data and develop application using them with Google Fit SDK. Google is

making efforts to create an environment where they access much information and develop good applications focusing on use of personal fitness data. Google Fit, in particular, provides the foundation upon which application and devices meet, giving birth to lots of new health management services depending on combinations of existing applications and devices.

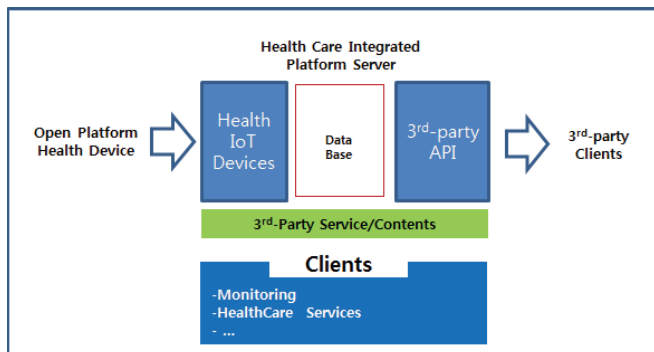
The digital healthcare industry is in close relations with health equipment and application that collect personal biometric information, platforms that store and manage the collected information, and the services using them. With accumulated development technology for wearable devices, generalized technology to store and analyze big data on health and bearable cost to establish them, the health management paradigm shifted from treatment to prevention with health management service converging medical business and IT. A variety of health devices were rapidly spread and products were released based on many development platforms accordingly. Nevertheless, there are many development platforms that are preferred per function because of the features that each development platform has. Even so, chances are high that the user will use multiple products developed from different platforms. In this paper, we suggested the development result of service and contents oriented open health management platform that comprehensively manages healthcare products created in many development platforms in a common ecosystem to build a personal health management system.

## 2 Open Healthcare Platform

The open healthcare platform suggested by the study is operated in close relation with platform operator, external service provider, client for health management and provision of health management device in and around the healthcare server.

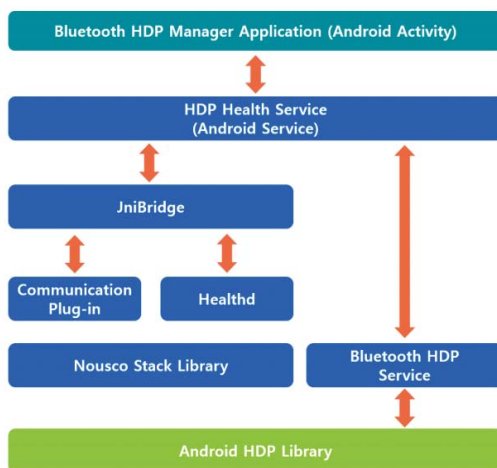
The health management platform server consists of platform administrator function, health device and linked API, PHI management website as well as health information big data analysis and inference engine that analyzes health information and provides service to the client as Fig.1. It also uses HTML 5-based template and verified components to offer development productivity and fancy UX. The algorithm was developed that can effectively distribute the increasingly

large volume of traffic along with increasing clients in the process of server establishment.



[Fig. 1] Open Health Care Platform Components

When the client applies the wearable health device produced with external open platforms like Google Fit or Samsung SAMI to this platform, the app client software for web and mobile supported by the health management platform was developed to enable the user to check his health condition and receive health management service and contents real time accordingly. In this research, we also developed android 4.x-based application so that the health management service and contents provided by a third party can be checked with the client PC as well as health management service and contents offered by web-based client application and third party can be operated in the mobile environment such as smartphone or tablet as Fig. 2.

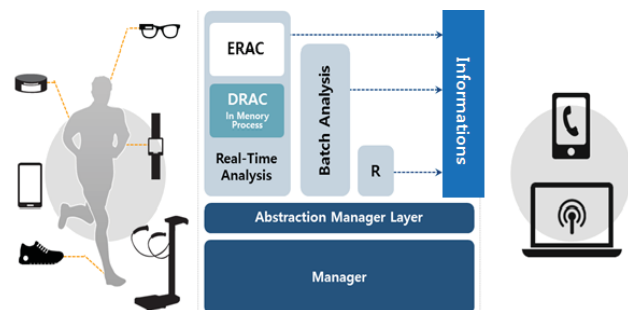


[Fig. 2] App Client Software Architecture on Android

The health management platform provides the means that can expand the function that the platform does not have in the form of open API (restful API) where third party solution or library can be linked to support API for the third party. It is composed to run as a Client-Server model. It sends the data requested by third party solution or library through API. The platform provides API in the form that enables calling through

https and communicates by writing the full text with JSON as the standard. Since access is possible with various platforms, it takes the form that can process data requested from any platform. Lastly, API was designed to have usable levels to allow discrimination on functions depending on the level of user authority. The session information was made to minimize loading against the server by having the client manage it. Cache function was also installed on the server to respond to the substantial requests at optimal speed, thus overcoming the performance issue.

The “health information big data analysis engine” embedded in the health management platform patterns and analyzes the feature of large volume of much health information generated or detected by the external health management device. After this, it induces useful information or rules and provides reliable health management information for the user based on objective data and description.



[Fig. 3] Health Information Big-data Analysis Engine

### 3 Conclusion

Unlike the developer-oriented healthcare platforms such as Google Fit, Samsung SAMI and Apple Health Kit, the platform in this research is user-oriented and has limited use to healthcare products and software. However, it can be easily connected to existing products and is compatible with products of different platforms.

### 4 References

- [1] Apple’s HealthKit. <https://developer.apple.com/healthkit>
- [2] Google Fit. <https://developers.google.com/fit/>
- [3] SamSund SAMI. <https://developer.samsungami.io/>
- [4] Flamingo Project. <http://sourceforge.net/projects/>

### Acknowledgment

This work (Grants No. C0275456) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2015.

