EJBI 2018 ISSN 1801 - 5603

# European Journal for Biomedical Informatics

Volume 14 (2018), Issue 1

www.ejbi.org

An Official Journal of the European Federation for Medical Informatics

# East Meets West

**Tomohiro Kuroda***

Division of Medical Information Technology and Administrative Planning, Kyoto University Hospital, Kyoto, Japan

**Correspondence to:**

**Professor Tomohiro Kuroda**
Division of Medical Information Technology and Administrative Planning ,
Kyoto University Hospital, Kyoto, Japan.
E-mail: tomo@kuhp.kyoto-u.ac.jp

**EJBI 2018; 14(1):01**
Received: February 13, 2018
Accepted: February 14, 2018
Published: February 21, 2018

Healthcare is both universal and local. Although the challenges can be easily shared globally, the solution must be strongly affected by local healthcare policy, social acceptance of general publics of the nation, and other various local factors. The other way around, a silver bullet for a very local problem of a small section of hospital can be found from the other side of the glove.

While working as a reviewer, the author came across so many reinventions of the wheels and too-simple opinions simply introducing cases of other culture as solutions of certain challenges without any modification (a person to express such opinions called "DEWANOKAMI" in Japanese). Most of the readers may have same experiences, I believe. The only way to avoid publishing such papers and opinions is to introduce reviewers and editors from other cultures.

The editorial committee of the European Journal of Biomedical Informatics (EJBI) newly invites three editors including the author from the board of the Japanese Association of Medical Informatics (JAMI) [1]. JAMI will encourage its members to submit papers and case reports to EJBI. We, Japanese editors, hope to make EJBI as a window to share insights and experiences between Japanese and European researchers.

One feature of Japanese medical informatics research is the existence of professors of university hospitals. As a result, many researches presented in JAMI annual conference series are strongly related with real clinical needs; some of them are about development and introduction of innovative hospital information systems and others are about data analysis for better hospital management.

For example, the author is the professor of medical informatics in school of medicine and of informatics, in addition to be the chief information officer of university hospital. The author designs, introduces, and manages a hospital information system, as well as conducts education and research on medical informatics for medical and informatics students. Additionally, the author is Ph.D. of computer science graduated an engineering course. It means that university students conduct medical informatics researches strongly related with needs of clinical field under computer scientist's direction. As a result, some of the prototypes developed as a Ph.D. or a master project are introduced to daily clinical activities.

The six papers from Japan published in this issue are typical examples.

I hope, European researchers will get some insights to promote their own researches from the papers and papers on comming issues.

Through his experience worked as visiting professor of mobile computing and augmented reality at the University of Oulu, the north end of Europe, in his early research career, the author knows how collaboration between strong European theoretical researches and strong Japanese development harvests wonderful academic fruit. The author expects that this newly opened window between Europe and Japan initializes lots of fruitful collaborations.

## References

1. Japanese Association of Medical Informatics : http://jami.jp/english/

# Computational Neuroscience

**Oleg Yu Mayorov**[1,2,3*]

[1] Kharkiv Medical Academy of Postgraduate Education, Ministry of Health of the Ukraine, Ukraine

[2] Institute for Medical Informatics and Telemedicine, Kharkiv, Ukraine

[3] Research Institute for Children and Adolescents Health Protection of the National Academy of Medical Sciences, Kharkiv, Ukraine

**Correspondence to:**

**Professor Oleg Yu. Mayorov,**
MD, PhD, Dr.Sci., Institute for Medical Informatics and Telemedicine, Kharkiv, Ukraine.
E-mail: o.y.mayorov@gmail.com

**EJBI 2018; 14(1):2-3**
Received: November 27, 2017
Accepted: December 12, 2017
published: January 10, 2018

One of the key components of computer neuroscience is the method of electroencephalography (EEG), which is widely used to study the activity of the human and animal's brain.

With the advent of the EEG, after a brief euphoria associated with the hope that this method will quickly reveal the secrets of the brain, an understanding emerged of the need to create and apply methods which can make a quantitative objective analysis of the electroencephalogram.

Quantitative EEG (qEEG) is an objective method of studying the central and peripheral nervous system, which is widely used in the clinic of internal diseases, neurology and psychiatry.

The qEEG significantly changed the classical EEG examination and gave qualitatively new possibilities for a quantitative assessment of the state of the nervous system.

This technology is developing very fast thanks to the achievements of microelectronics, computer technologies, communication technologies, mathematical methods of analysis and clinical informatics.

The effective use of encephalography for brain research and diagnosis, in the first place, depends on the qualitative recording of EEG signals. This is due to the understanding of the operation of hardware and software for recording EEG, compliance with the conditions in which the record is made in accordance with international standards. If these conditions are not met, the application of complex, expensive methods of mathematical analysis will not give the expected result. At the stage of EEG analysis it is necessary to choose one or several methods of analysis correctly depending on the task.

This requires a deep knowledge of the mathematical essence of the methods of analysis and capabilities available to the researcher of information diagnostic technologies for EEG analysis.

The most difficult stage is the neurophysiological and clinical interpretation, obtained as a result of the analysis of secondary mathematical indicators. This requires a deep knowledge of neuroanatomy, neurophysiology, and projection (extrapolation) of this knowledge with the neurological and psychiatric clinic.

One of the most sought-after areas is the algorithms of online and off-line detection of epileptiform activity for outpatient and 24-hour monitoring of patients. In addition, the development of a complex mathematical method of dipolar localization or the „solution of the inverse problem" makes it possible to establish a focus of epileptiform activity that is localized in the depth of the brain in subcortical structures, especially in the structures of the limbic system, which in some cases trigger epileptic seizures. There is a widespread introduction of EEG video monitoring to detect transient epileptiform activity and other transient neurological disorders. Video monitoring of EEG is also used to evaluate data for surgical treatment of seizures, and how intraoperative and resuscitative monitoring is in close contact with anesthesiology and neurosurgery.

Effective is the use of qEEG in sleep disorders for the recognition of sleep patterns - polysomnography, which is used to correctly diagnose and target treatment.

qEEG is used in newborns pathologies, pathological development of children and adolescents. The methods of qEEG are also used in the diagnosis of various clinical disorders: metabolic imbalance, endocrine disorders, toxic effects, coma, specific effects of drugs, and detection of brain death. The qEEG is used in conjunction with other methods of neuroimaging.

Neuropathologists and psychiatrists have a special need for methods of effective quantitative estimation of EEG for diagnosis and evaluation of the effectiveness of treatment of functional disorders - neuroses, psychoses, schizophrenia, when methods of medical visualization (tomography) are not effective.

As the power of personal computers increased, it became possible to implement complex mathematical algorithms and create new diagnostic technologies on their basis.

Widely used in the analysis of EEG methods of correlation, spectral and pair coherent analysis are not suitable for the study of cognitive functions of the human and animal brain. These functions are violated in children and require control over treatment and recovery - trauma, genetic disorders and congenital malformations.

There is an increase in the number of depressive and anxiety-phobic disorders, including children and, especially, adolescents. To study the objective quantitative indicators of cognitive impairment,

high efficiency is provided by the use of technologies based on multi-dimensional linear spectral analysis of the EEG.

However, the multidimensional linear analysis is the method for the analysis of linear systems and also the EEG considers as stochastic process. So, that does not allow evaluating comprehensively all EEG properties. Alternative to traditional approach is the hypothesis, according to which an EEG signal is necessary to consider as an output of non-linear system that is in a state of dynamic chaos.

For diagnostics of a condition of the brain or various mental diseases outcomes of the nonlinear analysis are extremely important. They - are objective and give the additional information. The basic information consists in evaluation of some "norm of chaos" ("norm of chaotic state"), and also in evaluation of deviations from "norm" (in this or that leg) under influence of those or other actions. It may be various mental, endocrine, neurological diseases, emotional stresses, which call deviations of a degree of a chaos (chaotic state) from "norm". Thus deviations in those and other legs may mean "illness".

If we will manage to reveal for given neurodynamic system (or an aspect of cerebral activity, for example, perception, memory, emotions, motivations etc.) for this or that cerebral process "norm" of a chaos deviations in both directions can be surveyed as "illness" and, hence, as a degradation. Further it is possible to inspect a choice of a technique of "treatment". Here the criterions of relative degree of chaotic state again enter in game. If by this criterion "treatment" approximates a condition of an open system to "norm", it means that process of self-organizing takes place. Otherwise "treatment" calls the further degradation. To solve the tasks described above the theoretical approach worked out, calculations algorithms, new secondary EEG (ESCoG) indices and their physiological interpretation are proposed to use.

The need to use sophisticated mathematical methods for EEG analysis for experimental and clinical studies today comes to the fore problems of standardization of methods, equipment and ongoing training of specialists - clinical neurophysiologists and technical personnel.

# Computational Ensemble Approach for Immune System Study: Conformational B-cell Epitope Prediction

**Yuh-Jyh Hu[1,2*], Shun-Ning You[1] and Chu-Ling Ko[1]**

[1] Department of Computer Science, National Chiao Tung University, 1001 University Rd, Hsinchu, Taiwan
[2] Institute of Biomedical Engineering, National Chiao Tung University, 1001 University Rd, Hsinchu, Taiwan

## Abstract

Various tools have been developed to predict B-cell epitopes. We proposed a multistrategy approach by integrating two ensemble learning techniques, namely bagging and meta-decision tree, with a threshold-based cost-sensitive method. By exploiting the synergy among multiple retrainable inductive learners, it directly learns a tree-like classification architecture from the data, and is not limited by a prespecified structure. In addition, we introduced a new three-dimensional sphere-based structural feature to improve the window-based linear features for increased residue description. We performed independent and cross-validation tests, and compared with previous ensemble meta-learners and state-of-the-art B-cell epitope prediction tools using bound-state and unbound-state antigens. The results demonstrated the superior performance of the bagging meta-decision tree approach compared with single epitope predictors, and showed performance comparable to previous meta-learners. The new approach—requiring no predictions from other B-cell epitope tools—is more flexible and applicable than are previous meta-learners relying on specific pretrained B-cell epitope prediction tools.

## Keywords

B-cell; Epitope; Meta-decision tree; Ensemble learning

**Correspondence to:**

**Yuh-Jyh Hu**
Department of Computer Science, National Chiao Tung University,
1001 University Rd, Hsinchu, Taiwan.
E-mail: yhu@cs.nctu.edu.tw

**EJBI 2017; 14(1):04-15**
received: September 01, 2017
accepted: October 26, 2017
published: November 06, 2017

## 1 Introduction

B-cell epitopes are specific regions on proteins recognized as antigen-binding sites by the antibodies of B cells. A detailed understanding of the interaction between antibodies and epitopes facilitates the development of diagnostics and therapeutics as well as rational design of preventive vaccines [1, 2, 3]. Therefore, generation of potent antibodies through reverse immunological approaches requires precise knowledge of epitopes. According to their structure and interaction with antibodies, epitopes can be classified as conformational and linear epitopes. A linear epitope is formed by a continuous sequence of amino acids, whereas a conformational epitope comprises discontinuous sections of the antigen's primary sequence; the discontinuous sections are close together in the three-dimensional (3D) space and interact with an antibody together. Approximately 10% of B-cell epitopes are linear, whereas the remaining 90% are conformational [4, 5].

The increasing availability of protein structures has facilitated the development of computational prediction tools by exploiting protein antigen structures. The following are some of the knowledge that has been used to elucidate these structures for epitope prediction: (a) spatial neighborhood information and a surface measure [6]; (b) local spatial context, accessible surface area (ASA) propensity and consolidated amino acid index [7]; and (c) the B-factor to detect atomic fluctuation [8]. Some studies have either adopted a hybrid approach combining structural and physicochemical features [9, 10], proposed a ensemble meta-learner incorporating consensus results from multiple prediction servers by using a voting mechanism [11], applied an ensemble of classifiers using various input features [12], or employed meta-learning based on stacking and cascade generalization [13].

In this paper, we propose the use of a meta-decision tree (MDT) approach [14] for B-cell epitope prediction. The previous meta-learners have relied on the classifications of other B-cell epitope prediction tools, which are trained and may not be retrained conveniently by the user from new data; nevertheless, the current proposed meta-classifier is independent of these types of pretrained B-cell epitope prediction servers. Consequently, our approach is more applicable and flexible than the previously developed meta-

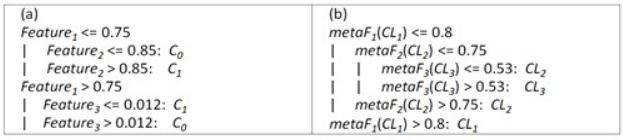
learning methods for B-cell epitope prediction. Our goal is not to develop a specific classifier for B-cell epitope prediction, but instead we intend to advocate the applicability of a generic improved ensemble meta learning approach over current B-cell epitope classifiers, and to show its competitive performance with those of other methods. To evaluate the new B-cell epitope meta-learner, we performed cross-validation (CV) with previous meta-classifiers and compared them with major epitope predictors for the same test data sets used previously for consistency. The results indicate that the proposed MDT approach outperforms commonly used single prediction servers considerably, as well as exhibiting performance comparable to previously developed meta-learners.

## 2 Materials and Methods

### 2.1 MDTs

MDTs [14] are used for meta-learning that applies multiple base classifiers to a single data set by exploiting the classification results of the base classifiers as a type of meta-knowledge. The structure of an MDT is identical to that of an ordinary decision tree, in that both have internal nodes and leaves, and have the same computational complexity $O$ ($m\ n\ log\ n$) where m is the number of attributes and n is the number of examples [15]; however, in an MDT, the attributes associated with the internal nodes and the meaning indicated by the leaves differ from those of an ordinary decision tree.

In both MDT and ordinary decision trees, an internal node specifies a test on an attribute value. For an ordinary decision tree, the attribute selected for the internal node must be one of the base attributes used to describe the data instances, for instance, the hydrophilic scale. By contrast, the attribute at an internal node of an MDT is a meta-attribute derived from the output of the base classifiers. Notably, although the base classifiers used in an MDT are standard inductive learners (e.g., artificial neutral network and naïve Bayes classifier), they differ from the B-cell epitope prediction servers (e.g., SEPPA 2.0 [7] and DiscoTope 2.0 [16] ) used as the base predictors by other meta-learning methods [11, 13]. Unlike these servers, the base classifiers used in MDTs can be retrained from new training data if required. As for the leaves, a leaf of an ordinary decision tree corresponds to a predicted class, whereas that of an MDT specifies a particular base classifier for class prediction. Figure 1 illustrates examples of an ordinary decision tree and an MDT. The ordinary decision tree in Figure 1A includes three internal nodes and four leaf nodes; each internal node specifies a test on a particular base attribute value [e.g., $Feature_1 \leq 0.75$ (or $> 0.75$)], and each leaf indicates the predicted class (e.g., $C_1$). The MDT in Figure 1B also has three internal nodes and four leaf nodes; unlike in the ordinary decision tree, each internal node in this MDT specifies a test on a particular meta-attribute derived from the output of




(A) An ordinary decision tree, (B) A meta decision tree.
Figure 1: Sample ordinary decision tree and MDT.

a base classifier [e.g., $metaF_1(CL_1)$ in Fig 1B], and rather than predicting the class, each leaf node predicts the base classifier most suitable for classification (e.g., $CL_1$). The advantage of an MDT is that it combines and exploits the classifications from multiple base classifiers to improve the accuracy of the final prediction.

### 2.2 Base Attributes and Meta-attributes

In the framework of machine learning, we translate an epitope prediction problem into an inductive learning problem. With a given training set of antigens with known epitope and nonepitope regions, the goal is to learn a classifier from a training set of antigens and apply the learned classifier to novel antigens for epitope detection.

To describe each amino acid on a protein antigen, we first adopted 14 base attributes: epitope propensity, secondary structure, residue accessibility, B factor, solvent-excluded surfaces, solvent-accessible surfaces, protein chain flexibility, hydrophilicity, position-specific scoring matrix (PSSM), atom volume, accessible surface area, side chain polarity, hydropathy index, and antigenic propensity. Descriptions and references of the base attributes have been reported by Hu et al. [13]. In addition, we considered five other physicochemical properties: surface probability [17], turns [18], exposed surface [19], and two types of polarities defined by Ponnuswamy et al. [20] and Grantham [21]. All the attributes were derived from the protein sequences or from the structural information provided by PDB. To keep the consistency, we prepared the attributes and their values for the training and testing on MDT in the same way without any discrepancy. Furthermore, we extended the idea of Ansari and Raghava [22] and created the 3D sphere-based attributes from the base attributes; the authors only considered amino acids in a linear window to generate one-dimensional (1D) window-based attributes. By contrast, we considered the amino acids in a constrained 3D spherical space and analyzed the values of different base attributes. To generate a 3D-based attribute value for an amino acid, we first created a surrounding sphere with its central carbon atom $C_\alpha$ as the center. We then computed the average value for a base attribute of all the amino acids within this 3D sphere and used the average as the 3D sphere-based attribute value for the center amino acid. Figure 2 illustrates a 3D spherical neighborhood for an amino acid. Because 90%

of the B-cell epitopes are conformational (discontinuous), 3D sphere-based attributes are more informative than 1D window-based attributes. By varying the radius of the sphere, we could define different 3D sphere-based attributes. With the base and 3D sphere-based attributes, we represented the protein antigens in the training set and trained the base classifiers from the base training data set.

A meta-attribute is defined over the output of the trained base classifiers. We used RFs [23], SVMs [24], C4.5 [25], k-NN [26], PART [27], BN [28], JRip [29], and VP [30] as the base classifiers. Furthermore, the majority vote of the base classifiers was also included as a base classification. According to Todorovski and Dzeroski [14], we calculated the properties of the class probability distributions predicted by the base classifiers, reflecting the certainty and confidence of the predictions. Here, we defined three meta-attributes: *epi_prob*(*x*,*B*), *entropy*(*x*,*B*), and *vote_epi_prop*(*x*), where *x* is a data instance and *B* a base classifier. The meta-attribute *epi_prob*(*x*,*B*) is the probability of epitope predicted by the base classifier *B* for the amino acid *x*. The meta-attribute *entropy*(*x*,*B*) is the entropy of the class probability distribution predicted by the base classifier *B* for the amino acid *x*. The meta-attribute *vote_epi_prop*(*x*) is the proportion of the epitope class predicted by all base classifiers for the amino acid *x*. These meta-attributes reflect the certainty of the base classifier in predicting the class, and they characterize the confidence variedly.

We computed the meta-attribute values for each data instance, namely the amino acid, on the basis of the output of the base classifiers and combined them to form a meta training data set. We then trained an MDT from a training set of data described by the meta-attributes. The metadata preparation process is illustrated in Figure 3. The meta training data can be obtained offline and independently of MDT, and consequently do not affect the training of MDT directly.

## 2.3 MDT Construction

MDT construction is identical to that of an ordinary decision tree. It involves a greedy, top-down, recursive search for the most suitable decision tree from a training data set. The core algorithm selects the most suitable attribute for an internal node and partitions the data available at the node into subsets according to the attribute values to create the descending nodes. This process is repeated for the data associated with each descendant to select the next attribute to grow the tree until some stopping criterion is satisfied.

Rather than employing the measures of impurity reduction commonly used for ordinary decision trees, such as information gain, gain ratio [25], and Gini [31], the focus

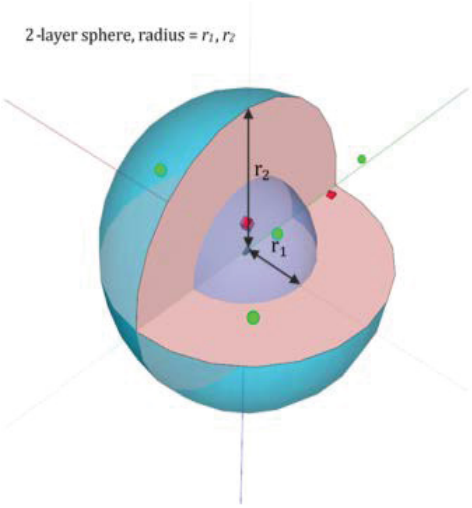



Figure 2: Two-layer 3D spherical neighborhood. A 3D sphere-based structural feature is defined for an amino acid (AA) based on its 3D spherical neighborhood specified by radius, *r*. The center of the sphere is the $C_\alpha$ of AA, and every other AA whose C is within the distance of *r* is considered a neighbor of AA. By varying *r*, we can specify different 3D spherical neighborhoods to define different 3D sphere-based structural features.

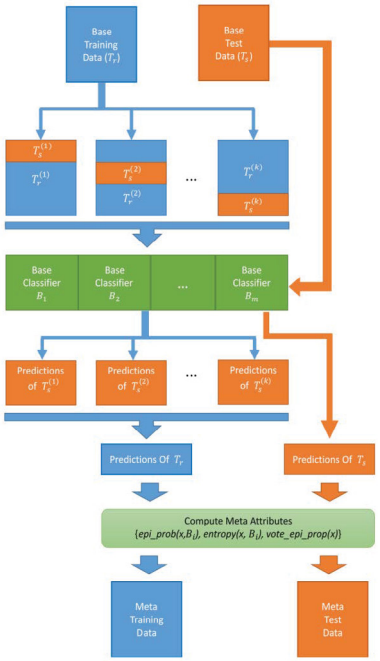



Figure 3: Metadata preparation process.Base training data *Tr* are divided into *k* folds. In each iteration, (*k*-1) folds of *Tr* are fed to all the base classifiers for training and the trained base classifiers are tested on the remaining fold to obtain the predictions. After repeating the same procedure *k* times, the predictions of the entire *Tr* are obtained from all base classifiers. To prepare meta training data, for each data instance *x*, the values of its three meta-attributes, *epi_prob*(*x*,$B_i$), *entropy*(*x*,$B_i$), and *vote_epi_prop*(*x*), are computed; these values together form the meta training data. Meta test data are prepared similarly.

of MDTs is the accuracy of each base classifier in predicting the data instance *S* available at an internal node. We defined the new information measure as follows:

$$info(S) = 1 - max_{B \in Base}\ Accuracy(B, S) \quad (1)$$

Where, *B* is a base classifier, *Base* is the set of all base classifiers, *S* is the data available at an internal node, and *Accuracy* is the classification accuracy of *B* on *S*. We selected the attribute that maximized the decrease in *info* of the subsets of *S* after the partition according to the values of the selected attribute compared with the original *info* of *S*. We repeated the same selection–partition process to grow the tree until the accuracy of some base classifier was 100% on the current subset or the size of the current subset was lower than a prespecified threshold. The classifier at the leaf node with the maximum accuracy was used to predict new instances after the tree grew completely.

## 2.4 Bagging and Decision Threshold

Three factors affect the performance of a predictor: (a) training d1ata, (b) representation bias, and (c) search bias. First, learning algorithms for predictors may produce different hypotheses from a small data set, and each has the same prediction accuracy on the same training data. Nevertheless, no single hypothesis can cover the entirety, or a sufficiently large portion, of the hypothesis space. Consequently, the overfitting problem can arise. Second, in most real-world applications, the true target concept may not be represented by any single hypothesis in the hypothesis space because of the limitations of representations, constraining the applicability of the learned hypothesis. Third, most learning algorithms adopt a local search strategy to prevent computational explosion during learning (e.g. the commonly used greedy partition rule for growing a decision tree); however, greedy local search of any form causes higher variance and can be stuck in local optima, thus failing to identify the true target.

An approach applicable for mitigating the aforementioned problems is ensemble learning. Various forms of ensemble learning have been developed. To further enhance the immunity of MDTs to the preceding problems, we applied a bagging-like strategy [32] to MDTs. By constructing an MDT ensemble from multiple random samples, we assumed that the final prediction based on the predictions of the MDTs ensemble can further reduce the variance among different MDTs, providing a more accurate approximation to the true target.

Most learning algorithms assume balanced class distributions and equal misclassification costs, which limit their applicability to epitope prediction because the B-cell epitopes are severely underrepresented. We adopted the undersampling strategy to decrease the number of nonepitopes in the training data to mitigate the class imbalance problem with B-cell epitope prediction. The appropriate epitope-to-nonepitope ratio for undersampling was first determined by either CV on training data or prior knowledge of class distribution in test data. We then performed multiple undersampling runs according to the selected ratio to create multiple random data sets to train an MDT ensemble.

We computed the probability for an amino acid of being epitope or nonepitope on the basis of the predictions of the MDTs trained from the random samples. By using *m* MDTs, we defined the scores of epitope, $Score_E$, and nonepitope, $Score_N$, for an amino acid *AA* as follows:

$$Score_E = \alpha \cdot \sum_{i=1}^{m} w_i \cdot \lceil e_i - 0.5 \rceil + (1-\alpha) \cdot \sum_{i=1}^{m} w_i \cdot e_i \quad (2)$$

$$Score_N = \alpha \cdot \sum_{i=1}^{m} w_i \cdot \lceil n_i - 0.5 \rceil + (1-\alpha) \cdot \sum_{i=1}^{m} w_i \cdot n_i \quad (3)$$

In equations (2) and (3), $e_i$ (or $n_i$) is the probability of being an epitope (or nonepitope) according to the prediction of the *i*th MDT. To predict the class of *AA*, we traversed each MDT to a leaf that specified a base classifier to make the prediction. We obtained the proportion of a base classifier used to make predictions from the *m* MDTs and denoted the proportion for each base classifier by using $w_i$. A higher $w_i$ value indicates a stronger weight of that base classifier exerted on the score; when $w_i$ is set to 1, all the base classifiers are treated equally. The first term in equations (2) and (3) considers only the count of classifications by the *m* MDTs, whereas the second term considers the class probabilities. We used a control parameter $\alpha$ to balance the effects of the two scoring mechanisms, and its value could be determined through CV. We defined the probability for the amino acid *AA* of being epitope or nonepitope as follows:

$$P_N(AA) = \frac{Score_N}{Score_E + Score_N} \quad (4)$$

$$P_N(AA) = \frac{Score_N}{Score_E + Score_N} \quad (5)$$

To appropriately address the imbalanced class distribution in B-cell epitopes, we also set a probability threshold for the final classification as follows:

$$Class(AA) = \begin{cases} non-epitope, & P_E(AA) < \theta \\ epitope, & P_E(AA) \geq \theta \end{cases} \quad (6)$$

Where $\theta$ is a threshold. A carefully selected on the basis of CV or prior knowledge warrants a reasonable performance of the class-sensitive bagging MDT approach. Figure 4 illustrates the entire flow of this system.

## 2.5 Data Sets and Performance Measures

We collected the training and test data used in DiscoTope 2.0 [16], SEPPA 2.0 [7], Bpredictor [33], ElliPro [34], CBTOPE [22], EPMeta [10], and B-cell meta-classifiers [13] and by Zhao et al. [35]; we then combined these with the data in the Epitome database [36] and Immune Epitope Database

(IEDB) [37] to prepare the data set for the comparative study. After removing the duplicates and filtering out the antigens without epitope residues annotated in either Epitope Information or B cell Assay Information in the IEDB, we obtained a total of 363 bound-state antigens. Because the epitope predictors used in our study were web-based servers or software packages that could not be retrained using different training data, to conduct a consistent and unbiased comparative analysis of the prediction performances of these predictors, we created an independent data set of antigens with known epitopes. We divided this data set of 363 protein antigens into a test data set and a training data set. To ensure fair comparison between different prediction methods, we selected 18 antigens that were not used before to train any of the predictors in comparison for testing, and used the remaining 345 antigens that were used previously to train these predictors for MDT training. Table 1 lists the total 363 antigens, and Table 2 lists the 18 test antigens.

While most of the studies of epitope prediction and feature analysis are focused on bound-state antigen structures [6, 7, 16, 34, 35], epitopes in bound states show different characteristics and reveal more binding information than unbound epitopes [38], which can raise two issues in epitope prediction. One is that explicit binding information in the bound-state structures can bias the prediction performance; the second is that an antigen possibly bound by multiple antibodies can cause more false negatives because only the epitope to the antibody in the bound structure is considered a true epitope, and all remaining epitopes to other antibodies are labeled as nonepitopes. We adopted a set of unbound-state antigens, listed in Table 3, recently constructed and annotated by Ren et al. [39] to evaluate MDT's performance for unbound-state epitope prediction, and compare it with other epitope predictors.

We evaluated prediction performances by using several measures: TPR (i.e., sensitivity), FPR, precision (i.e., positive predictive value), percentage accuracy, F-score, and MCC. Table 4 lists the definitions of these measures. In general, correlation exists between the TPR and FPR produced by a predictor. Typically, the FPR increases with the TPR. We prepared ROC curves to summarize the results on the different thresholds.

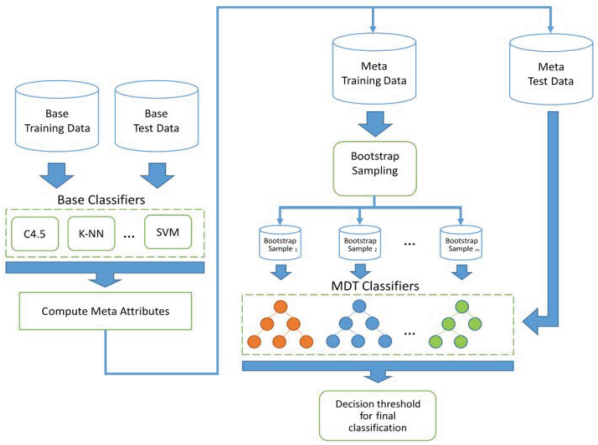



Figure 4: System flow of bagging MDT.

## 2.6 Correlation Analysis of Base Classifiers and Ablation Analysis of Base Attributes

An MDT can be constructed from an arbitrary number of different base classifiers, and its overall performance depends on these learning components. If the learning components have complementary predictive strengths, an MDT can search various hypotheses in the hypothesis space and provide superior generalizations for novel test data to those of a single-component learner. We used the ARI [40] to measure the strength of the relationship between the predictions produced by two base classifiers. Although the ARI was initially designed to measure agreement between two clustering results, in our case, a higher ARI value could indicate greater agreement between the two classifiers. If $P$ is the partition of the amino acids into epitopes and nonepitopes for a given data set of antigens, according to the predictions of the classifier $A$, and $Q$ is the partition produced by the classifier $B$, a lower ARI value between $P$ and $Q$ suggests a higher probability that the two classifiers have complementary strengths. After evaluating several different indices for the measurement of the agreement between two partitions, Milligan and Cooper [41] recommended the use of the ARI. Therefore, in this study, we adopted the ARI for evaluating the correlation between the classifiers. The results from the ARI analysis provided a basis for selecting the appropriate base classifiers in MDTs.

In addition to assessing the complementary prediction strengths of the base classifiers by using the ARI, we conducted an ablation analysis of different base attribute types to measure their contribution to the MDT. We classified the base attributes into three categories: (a) sequence, (b) structure, and (c) 3D sphere-based. We compared the relevance of the three categories of base attributes to MDT by their removal or addition and estimated their effects on the meta-classification by the amount of decrease or increase in prediction performance.

# 3 Results

To keep the consistency in evaluating MDT's performances, we followed the same protocol to prepare the attributes and their values for MDT's training as well as its testing without any discrepancy. All the attributes were derived from the protein sequences or from the structural information provided by PDB.

## 3.1 Prediction Correlations Between Base Classifiers

For a meta-learning method to perform effectively, the base classifiers must have complementary predictive capabilities,

Table 1: Data set of 363 protein antigens.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1A2Y_C | 1ADQ_A | 1AFV_A | 1AHW_C | 1AHW_F | 1AR1_B | 1BGX_T | 1BJ1_V | 1BJ1_W | 1BQL_Y |
| 1BVK_C | 1BVK_F | 1BZQ_A | 1C08_C | 1CIC_C | 1CIC_D | 1CZ8_V | 1CZ8_W | 1DQJ_C | 1DVF_A |
| 1DVF_B | 1DZB_X | 1DZB_Y | 1E6J_P | 1EGJ_A | 1EO8_A | 1EZV_E | 1FBI_X | 1FDL_Y | 1FJ1_F |
| 1FNS_A | 1FSK_A | 1FSK_D | 1FSK_G | 1FSK_J | 1G7H_C | 1G7I_C | 1G7J_C | 1G7L_C | 1G7M_C |
| 1G9M_G | 1G9N_G | 1GC1_G | 1HYS_A | 1HYS_B | 1I9R_A | 1IAI_H | 1IAI_L | 1IC4_Y | 1IC5_Y |
| 1IC7_Y | 1J1O_Y | 1J1P_Y | 1J1X_Y | 1J5O_B | 1JHL_A | 1JPS_T | 1JRH_I | 1JTO_L | 1JTO_M |
| 1JTP_L | 1JTP_M | 1JTT_L | 1KB5_A | 1KB9_E | 1KEN_A | 1KIP_C | 1KIQ_C | 1KIR_C | 1KXQ_A |
| 1KXT_A | 1KXV_A | 1KYO_E | 1LK3_A | 1LK3_B | 1MEL_L | 1MEL_M | 1MHP_B | 1MLC_E | 1MLC_F |
| 1N5Y_B | 1N6Q_B | 1N8Z_C | 1NBY_C | 1NBZ_C | 1NDG_C | 1NDM_C | 1NSN_S | 1OAK_A | 1OAZ_A |
| 1OP9_B | 1ORQ_C | 1ORS_C | 1OSP_O | 1OTS_A | 1OTS_B | 1OTT_A | 1OTT_B | 1OTU_A | 1OTU_B |
| 1P2C_C | 1P84_E | 1PG7_L | 1PKQ_J | 1QFW_A | 1QLE_B | 1R0A_B | 1R3I_C | 1R3J_C | 1R3K_C |
| 1R3L_C | 1RI8_B | 1RJC_B | 1RJL_C | 1RVF_1 | 1RVF_2 | 1RVF_3 | 1RZJ_G | 1S78_B | 1SQ2_L |
| 1T6V_L | 1T6V_M | 1TPX_A | 1TQB_A | 1TQC_A | 1TZH_V | 1TZH_W | 1TZI_V | 1UA6_Y | 1UAC_Y |
| 1UJ3_C | 1V7M_V | 1VFB_C | 1W72_A | 1WEJ_F | 1XGP_C | 1XGQ_C | 1XGR_C | 1XGT_C | 1XGU_C |
| 1XIW_A | 1XIW_E | 1YJD_C | 1YQV_Y | 1YY9_A | 1YYM_G | 1Z3G_A | 1Z3G_B | 1ZA3_R | 1ZTX_E |
| 1ZV5_L | 2AEP_A | 2ARJ_Q | 2B2X_A | 2BDN_A | 2BOB_C | 2BOC_C | 2DD8_S | 2DQC_Y | 2DQD_Y |
| 2DQE_Y | 2DQF_C | 2DQF_F | 2DQG_Y | 2DQH_Y | 2DQI_Y | 2DQJ_Y | 2DWD_C | 2DWE_C | 2EIZ_C |
| 2EKS_C | 2FJG_V | 2FJG_W | 2GHW_A | 2H9G_R | 2H9G_S | 2HMI_B | 2I25_L | 2I25_M | 2I26_L |
| 2I26_M | 2I26_Q | 2I60_G | 2I60_P | 2I9L_I | 2IFF_Y | 2ITD_C | 2J4W_D | 2J5L_A | 2JEL_P |
| 2NXY_A | 2NXZ_A | 2NY0_A | 2NY1_A | 2NY2_A | 2NY3_A | 2NY4_A | 2NY5_G | 2NY6_A | 2NY7_G |
| 2OZ4_A | 2P42_A | 2P42_C | 2P43_A | 2P44_A | 2P45_A | 2P46_A | 2P46_C | 2P47_A | 2P48_A |
| 2P49_A | 2P4A_A | 2P4A_C | 2Q8A_A | 2Q8B_A | 2QQK_A | 2QQN_A | 2R29_A | 2R4R_A | 2R4S_A |
| 2R56_A | 2UZI_R | 2VH5_R | 2VIR_C | 2VIS_C | 2VIT_C | 2VXQ_A | 2VXS_A | 2VXS_B | 2VXT_I |
| 2W9E_A | 2XQB_A | 2XQY_A | 2XQY_E | 2XTJ_A | 2XWT_C | 2YBR_C | 2YBR_F | 2YBR_I | 2YC1_C |
| 2YC1_F | 2YSS_C | 2ZJS_Y | 2ZNW_Y | 2ZNW_Z | 2ZNX_Y | 2ZNX_Z | 2ZUQ_A | 3A67_Y | 3A6B_Y |
| 3A6C_Y | 3B9K_B | 3BDY_V | 3BE1_A | 3BGF_A | 3BGF_S | 3BQU_B | 3BSZ_E | 3BSZ_F | 3C09_A |
| 3CVH_A | 3CVH_M | 3D85_C | 3D9A_C | 3DVG_Y | 3DVN_V | 3EO1_C | 3EO1_F | 3FMG_A | 3G04_C |
| 3G6D_A | 3GB7_C | 3GBM_B | 3GBN_B | 3GI8_C | 3GI9_C | 3GRW_A | 3H42_A | 3H42_B | 3HFM_Y |
| 3HI6_A | 3HI6_B | 3HMX_A | 3I50_E | 3IDX_G | 3IGA_C | 3J1S_A | 3K3Q_B | 3K3Q_C | 3KJ4_A |
| 3KJ6_A | 3KR3_D | 3L5W_I | 3L5X_A | 3LD8_A | 3LDB_A | 3LEV_A | 3LH2_S | 3LH2_T | 3LH2_U |
| 3LH2_V | 3LHP_S | 3LHP_T | 3MA9_A | 3MJ9_A | 3MXW_A | 3NGB_G | 3NH7_A | 3O0R_B | 3O0R_C |
| 3O2D_A | 3PGF_A | 3Q3G_E | 3QA3_E | 3QA3_G | 3QA3_I | 3QA3_L | 3QWO_P | 3R1G_B | 3RU8_X |
| 3RVV_A | 3RVW_A | 3RVX_A | 3SDY_A | 3SDY_B | 3SE8_G | 3SE9_G | 3SKJ_E | 3SKJ_F | 3SOB_B |
| 3SQO_A | 3T2N_A | 3THM_F | 3TJE_F | 3U2S_C | 3U2S_G | 3U30_D | 3UC0_A | 3UC0_B | 3UX9_A |
| 3UX9_C | 3VG9_A | 4AEI_A | 4AEI_B | 4AEI_C | 4AL8_C | 4ALA_C | 4DGI_A | 4DKE_A | 4DKE_B |
| 4DKF_A | 4DKF_B | 4DN4_M | 4DTG_K | 4ETQ_C | 4F2M_E | 4F2M_F | 4F3F_C | 4FQI_B | 4GMS_A |
| 4GMS_C | 4GMS_E | 4HKX_E | | | | | | | |

Table 2: Independent test data set of 18 protein antigens.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1BZQ_A | 1KXT_A | 1KXV_A | 1W72_A | 2I9L_I | 2J4W_D | 2J5L_A | 2OZ4_A | 2R4R_A | 2R4S_A |
| 2ZJS_Y | 3B9K_B | 3BQU_B | 3BSZ_E | 3BSZ_F | 3DVN_V | 3KJ4_A | 3KJ6_A | | |

Table 3: Data set of unbound-state antigens.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3TGT | 4GXX | 4OIE | 2I5V | 1J95 | 2QTW | 1D7P | 1B1I | 1YG9 |
| 4I53 | 4M4Y | 2C36 | 2G7C | 1JVM | 1KF3 | 1F45 | 1DOK | 1FCQ |
| 4JPJ | 1F8D | 4OSN | 1POH | 1DKK | 1ZVM | 2ILK | 1KEX | 1BV1 |
| 4DKP | 3NN9 | 2GHV | 1UB4 | 1HHL | 3OIW | 2NVH | 3M1N | 3PX8 |
| 3O3X | 4NN9 | 3K7B | 1WHO | 1JSE | 4KZN | 1MF7 | 3MJ6 | |
| 4IPY | 3IRC | 4E9O | 1I4M | 2NWD | 4EFV | 1MJN | 3NCL | |
| 3NTE | 3VTT | 3EJC | 3Q27 | 2VB1 | 1BOY | 2ICA | 3S26 | |
| 1HGH | 3WE1 | 1W8K | 4NX7 | 1HX0 | 1ATZ | 3FCU | 4GNY | |
| 3KU3 | 3GGQ | 1Z40 | 3KVD | 3ZKG | 1AUQ | 2YXF | 1AHO | |
| 3ZP0 | 2HG0 | 3WKL | 1EY0 | 3Q6O | 1IJB | 3O1Y | 3D6S | |
| 4FNK | 2P5P | 2WK0 | 1IGD | 2GBC | 1BIO | 4JNI | 3F5V | |

Table 4: Definitions of performance measures.

| Performance Measure | Definition |
|---|---|
| TPR[a] | TP/(TP+FN) |
| FPR | FP/(FP+TN) |
| Precision[b] | TP/(TP+FP) |
| Accuracy | (TP+TN)/(TP+TN+FP+FN) |
| F-score | 2×TPR×Precision/(TPR+Precision) |
| MCC | $\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$ |

[a]TPR is also known as sensitivity or recall.
[b]Precision is also known as positive predictive value.

which can be reflected by relatively low correlation among their predictions. We selected eight base classifiers to construct an MDT. They are random forests (RFs) [23], support vector machines (SVMs) [24], C4.5 [25], k-nearest neighbors (k-NN) [26], PART [27], Bayes Net (BN) [28], JRip [29] and Voted Perceptron (VP) [30].

We measured the correlation between two base classifiers by the adjusted Rand index (ARI) [40] of their classifications. Table 5 lists the ARI values of all pairs of the base classifiers for an independent test data set of 18 antigens (see Materials and methods). The mean ± standard deviation ARI values for the test data sets were 0.238 ± 0.084; the ARI value is relatively low, indicating a relatively weak correlation among the base classifiers.

The 18-antigen independent test data set contained a total of 243 epitope and 3,760 nonepitope residues. A base classifier was trained using the training sets of 345 antigens; for the test set, it classified a protein residue as an epitope or a nonepitope. For each epitope and nonepitope residue, we counted the number of base tools that correctly classified the residue as epitope or nonepitope. Figure 5 presents the distributions of epitope and nonepitope residues for all test proteins, according to the number of base tools with the same prediction, for indicating the degree of agreement in classification among the base tools. For instance, as shown in Fig 5, we observed only 1.6% of the epitope residues were correctly and unanimously predicted by all base classifiers, whereas 32.5% of the epitopes could not be detected by any base predictor. By contrast, more than 65% of the epitope residues were correctly classified by 1–7 base predictors. Compared with epitopes, a markedly higher percentage of nonepitopes was correctly classified by all base predictors. Despite the higher percentage of unanimous predictions, more than 30% of the nonepitopes were classified variedly, and they were correctly predicted by 1–7 base classifiers. Taken together, these results indicate that base classifiers do not always agree when predicting epitopes and that they may have complementary strengths, suggesting that a meta-learner built based on these base learners can demonstrate synergy in their predictive capabilities.

## 3.2 Independent Test of MDT and other B-cell Epitope predictors

Five representative conformational epitope predictors [7, 16, 22, 33, 34], and three meta-classifiers [11, 13] were considered for comparison in the test. We compared MDT with current B-cell epitope predictors on a test data set of 18 antigens selected from a total of 363 bound antigens that were previously used to train and test epitope prediction tools (see Materials and methods). We did not perform k-fold cross-validation (CV) based on the 363 antigens because some epitope predictors, such as DiscoTope 2.0 and SEPPA 2.0 in this comparative study, had been pretrained using different antigens from the 363 antigens, and the antigens previously used for training can overlap the test data folds in the iterative process of CV. This violates the principles of CV, and is likely to cause overestimated performance. To avoid the bias, we performed an independent test instead.

We trained a bagging MDT model from 354 antigens, which were used previously to train the predictors in comparison, and compared its performance with the performances of other B-cell epitope predictors for the same test set of 18 antigens. The 18 test antigens were not used before to train any of the predictors in comparison to ensure a fair comparison. For each predictor in comparison, we selected the parameter values of their best-performing models for the training data set individually; these values were used in the tests to ensure fair comparison. Table 6 shows that the meta-classifiers, cascade, stacking, and bagging MDT, all considerably outperformed the five representative conformational epitope predictors for ACC, F-score, and MCC. The results revealed that DiscoTope 2.0 [16], ElliPro [34], and Bpredictor [33] presented high true positive rates (TPRs) of prediction; nevertheless, they also had high false positive rates (FPRs). By contrast, SEPPA 2.0 [7] and CBTOPE [22] showed lower TPRs as well as FPRs. Compared with these single predictors, the meta-classifiers, cascade, stacking, and bagging MDT led to a more favorable balance between TPRs and FPRs, and consequently demonstrated a higher F-score and MCC. Figure 6 shows their ROC curves, indicating the trade-off between the amounts of true positives (TP) and false positives (FP) produced by the classifiers. In general, these observations suggest that the performance of bagging MDT is superior to that of other current B-cell epitope prediction methods. In addition, the bagging MDT approach is comparable with previously reported epitope meta-classifiers, such as EPMeta [11], cascade, and stacking [13], while unlike previous meta-classifiers, it does not depend on the prediction output of specific pretrained B-cell epitope predictors such as DiscoTope 2.0 [16] and SEPPA 2.0 [7]. Notably, when bagging MDT also employed the output of the eight prediction tools, e.g. SEPPA 2.0, as used in cascade and stacking [13], it produced the highest performance with regard to ACC, F-score, and MCC, as presented in the final row of Table 6.

Table 5: Correlation analysis of base classifiers based on 18 antigens.

| 18 Ags | C4.5 | KNN | Voted Perceptron | PART | Random Forest | Bayes Net | JRip |
|---|---|---|---|---|---|---|---|
| **KNN** | 0.198 | - | - | - | - | - | - |
| **Voted Perceptron** | 0.189 | 0.256 | - | - | - | - | - |
| **PART** | 0.290 | 0.259 | 0.282 | - | - | - | - |
| **Random Forest** | 0.157 | 0.245 | 0.166 | 0.164 | - | - | - |
| **BayesNet** | 0.232 | 0.120 | 0.126 | 0.240 | 0.050 | - | - |
| **JRip** | 0.251 | 0.239 | 0.281 | 0.306 | 0.237 | 0.191 | - |
| **SVM** | 0.248 | 0.359 | 0.382 | 0.386 | 0.361 | 0.125 | 0.335 |

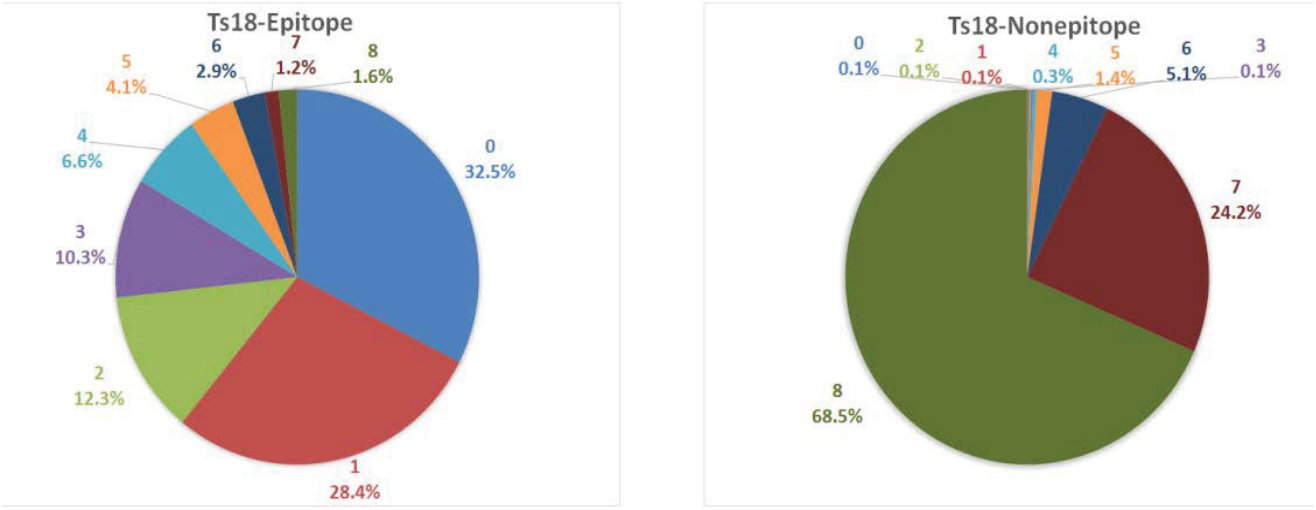



Figure 5: Pie charts of classification agreement. Pie charts showing the degree of agreement among base tools in epitope prediction of 18 test protein antigens. Distributions of the counts of (A) epitope and (B) nonepitope residues, according to the number of base tools with the same prediction.

Table 6: Results of independent test.

| Classifier | TPR | FPR | PPV | ACC | F-score | MCC |
|---|---|---|---|---|---|---|
| **SEPPA** | 0.140 | 0.047 | 0.162 | 0.904 | 0.150 | 0.100 |
| **DiscoTope** | **0.959** | 0.756 | 0.076 | 0.287 | 0.140 | 0.115 |
| **ElliPro** | 0.807 | 0.558 | 0.085 | 0.464 | 0.155 | 0.120 |
| **Bpredictor** | 0.856 | 0.634 | 0.080 | 0.396 | 0.147 | 0.111 |
| **CBTOPE** | 0.029 | **0.001** | 0.636 | **0.940** | 0.055 | 0.127 |
| **EPMETA** | 0.757 | 0.522 | 0.086 | 0.495 | 0.154 | 0.112 |
| **Cascade** | 0.222 | 0.016 | 0.466 | 0.937 | 0.301 | 0.293 |
| **Stacking** | 0.243 | 0.015 | **0.513** | **0.940** | **0.330** | **0.326** |
| **F19_BaggingMDT_BaseOnly**[a] | 0.148 | 0.017 | 0.356 | 0.932 | 0.209 | 0.199 |
| **F19_BaggingMDT_plus3Dsphere**[b] | 0.255 | 0.019 | 0.466 | 0.937 | **0.330** | 0.315 |
| ***F19_BaggingMDT_plus3Dsphere_8T***[c] | *0.272* | *0.014* | *0.555* | *0.943* | *0.365* | *0.362* |

[a]Bagging MDT used only 19 base attributes and no 3D sphere-based attributes.
[b]Bagging MDT used 19 base attributes and the derived 3D sphere-based attributes.
[c]Bagging MDT used 19 base attributes, the derived 3D sphere-based attributes, as well as the output of the eight prediction tools, e.g. SEPPA, used in Cascade and Stacking.

## 3.3 Individual Comparisons Between MDT and Other B-cell Epitope Predictors

In addition to the tests of bagging MDT and the commonly used epitope predictors on the same test data, we compared bagging MDT with the epitope predictors separately by using different data sets. We tested six representative epitope predictors: SEPPA 2.0 [7], DiscoTope 2.0 [16], ElliPro [34], Bpredictor [33], CBTOPE [22], and EPMeta [11]. All had been trained and tested by different data sets. In each test, we selected one epitope predictor

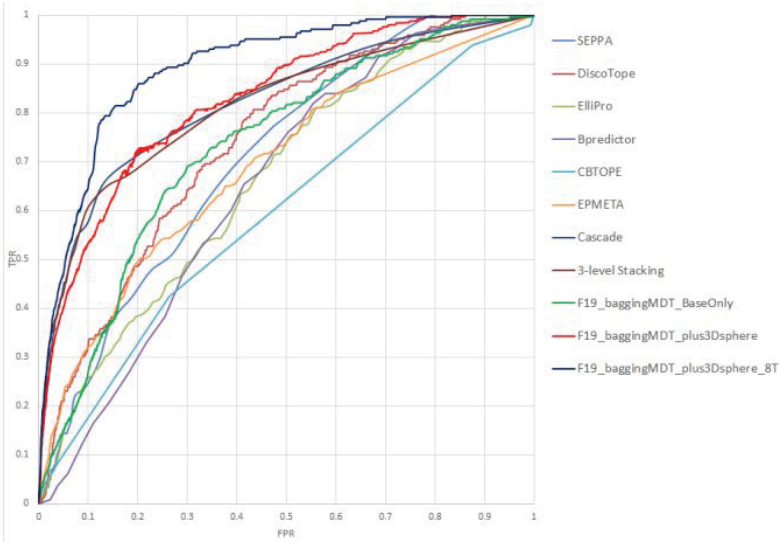



Figure 6: ROC curves of predictors. ROC curves of epitope predictors and meta-classifiers based on the test data set of 18 antigens.

for comparison. To make a fair and consistent comparison, we only trained and tested bagging MDT on the same data sets that had been used specifically to train and test the predictor selected for comparison. The results of the individual comparisons are presented in Table 7. Bagging MDT considerably outperformed SEPPA 2.0, ElliPro, Bpredictor, CBTOPE, and EPMeta in the individual tests except DiscoTope 2.0. As shown in Table 7, in the individual comparison with DiscoTope 2.0, bagging MDT was overconservative, as indicated by its significantly lower TPR and FPR, which caused weaker performance than DiscoTope 2.0. In general, these results demonstrate that the synergy in multiple base classifier effects in MDT can result in performance superior to other current epitope predictors. Furthermore, unlike previous meta-classifiers, the new bagging MDT approach does not require the prediction output of other epitope predictors [11] or a prespecified classification architecture [13], and is thus more flexible and applicable.

### 3.4 Evaluation of MDT based on Unbound-State Antigens

In addition to the independent test for bound-state antigens, we compared MDT with other epitope predictors by a set of unbound-state antigens. To maintain the consistency with the previous study in [39], we used the same unbound-state antigen data and also conducted antigen-based 10-fold CV by dividing the same unbound-state structures into 10 subsets randomly. The overall performance was used as the average of the results obtained from all iterations of three 10-fold CVs. The experimental results in Table 8 demonstrated that MDT outperformed the structure-based predictors DiscoTope 2.0, SEPPA 2.0 and ElliPro markedly for MCC, and showed comparable performance in F-score. Table 8 also showed that MDT was competitive with a two-stage SVM-based unbound epitope predictor PUPre [39]. In addition, the significantly higher precision and lower recall of MDT suggested that MDT was more conservative than the other tools for prediction of epitopes in unbound-state antigens.

### 3.5 Ablation Analysis

An ablation study provides insight into the effects of base learners on the prediction performance of a meta classifier. However, the time required for a complete ablation analysis increases exponentially with the number of base learners. To avoid computational explosion, following [13], we adopted a greedy approach for the ablation study. We used the same 94 antigens in [13] for training, and tested the trained classifiers on an independent set of 69 antigens.

We adopted two greedy iterative approaches, backward elimination and forward selection, to evaluate the contributions of available base classifiers. The greedy iterative backward elimination approach started with the maximal MDT built upon all available base learners. In each iteration, we identified the base classifier in MDT such that the performance of MDT decreased the most after the removal of this classifier. By contrast, the greedy iterative forward selection approach started with the minimal MDT built without any base classifier. In each iteration, we identified the base classifier available such that the addition of this classifier improved MDT's performance the most. We compared the relevance of the base classifier to MDT by the order of their removal or addition, and estimated their effects on MDT by the amount of decrease or increase in prediction performance. We show the results in Tables 9 and 10.

## 4 Conclusion

Numerous approaches can be used for predicting linear and conformational B-cell epitopes. Of these, the approach based

Table 7: Results of individual comparisons.

| Classifier | TPR | FPR | PPV | ACC | F-score | MCC |
|---|---|---|---|---|---|---|
| SEPPA[a] | 0.138 | 0.040 | 0.161 | 0.916 | 0.148 | 0.105 |
| F19_BaggingMDT_plus3Dsphere | 0.277 | 0.044 | 0.260 | **0.920** | **0.268** | **0.226** |
| DiscoTope[a] | 0.891 | 0.589 | 0.091 | 0.441 | **0.166** | **0.149** |
| F19_BaggingMDT_plus3Dsphere | 0.045 | 0.006 | 0.350 | **0.935** | 0.080 | 0.107 |
| ElliPro[b] | 0.740 | 0.519 | 0.125 | 0.505 | 0.214 | 0.128 |
| F19_BaggingMDT_plus3Dsphere | 0.606 | 0.025 | 0.707 | **0.941** | **0.653** | **0.623** |
| Bpredictor Ts=122[c] | 0.093 | 0.049 | 0.125 | 0.892 | 0.107 | 0.051 |
| F19_BaggingMDT_plus3Dsphere | 0.486 | 0.028 | 0.566 | **0.938** | **0.523** | **0.492** |
| CBTOPE[d] | 0.647 | 0.164 | 0.236 | 0.823 | 0.346 | 0.314 |
| F19_BaggingMDT_plus3Dsphere | 0.785 | 0.012 | 0.834 | **0.973** | **0.809** | **0.795** |
| EPMeta Ts=149[e] | 0.871 | 0.579 | 0.096 | 0.450 | 0.173 | 0.148 |
| F19_BaggingMDT_plus3Dsphere | 0.343 | 0.021 | 0.533 | **0.937** | **0.417** | **0.396** |

[a]Bagging MDT was trained and tested using the same data sets used specifically to train and test SEPPA 2.0 (or DiscoTope 2.0), excluding the antigens with missing feature values. All classifiers were tested on the same test data to conduct a consistent comparison.
[b]ElliPro only provided the test data but no training data. Consequently, we trained bagging MDT from our training data used in this study, excluding the antigens in ElliPro's test data set. We compared bagging MDT and ElliPro on the same test data previously used by ElliPro, excluding the antigens with missing feature values, to maintain consistency.
[c]Bagging MDT was trained on the same data previously used to train Bpredictor, excluding the antigens with missing feature values. Although Bpredictor provided its original test data set, the data lacked the epitope residues annotated in the IEDB. Alternatively, from the 363 antigens in this study we selected 122 antigens that were not used for Bpredictor training as the test data.
[d]Bagging MDT was trained and tested using the 60% nonredundant benchmark dataset previously used to evaluate CBTOPE, excluding the antigens with missing feature values. Following CBTOPE, we adopted 5-fold CV to compare the performances. We selected the parameter value of the best-performing CBTOPE on the training data set. All the classifiers, including CBTOPE, were tested on the same test data for consistent comparison.
[e]EPMeta only provided the training data but no test data. Consequently, we trained bagging MDT from the same training data used previously to train EPMeta, excluding the antigens with missing feature values, and performed a comparison between bagging MDT and EPMeta on the same test data set of 149 antigens selected from the 363 antigens in this study.

Table 8: Results of cross-validation on unbound-state antigens.

| Classifier | Recall | Precision | F-score | MCC |
|---|---|---|---|---|
| **SEPPA** | 0.48 | 0.16 | 0.24 | 0.14 |
| **DiscoTope** | 0.26 | 0.17 | 0.21 | 0.11 |
| **ElliPro** | 0.68 | 0.12 | 0.20 | 0.08 |
| **PUPre** | 0.71 | 0.18 | 0.28 | 0.21 |
| **F19_baggingMDT_plus3Dsphere[a]** | 0.13 | 0.53 | 0.21 | 0.23 |

[a]Bagging MDT used 19 base attributes and the derived 3D sphere-based attributes.

Table 9: Results of backward ablation analysis.

| MDT* | TPR | FPR | PPV | ACC | F-score | MCC |
|---|---|---|---|---|---|---|
| **F19_BaggingMDT_plus3Dsphere** | 0.463 | 0.011 | 0.791 | 0.947 | 0.584 | 0.581 |
| **\ SVM** | 0.483 | 0.024 | 0.633 | 0.936 | 0.548 | 0.520 |
| **\ Voted Perceptron** | 0.522 | 0.034 | 0.573 | 0.931 | 0.546 | 0.509 |
| **\ KNN** | 0.525 | 0.039 | 0.539 | 0.926 | 0.532 | 0.492 |
| **\ Random Forest** | 0.461 | 0.041 | 0.492 | 0.919 | 0.476 | 0.432 |
| **\ JRip** | 0.501 | 0.056 | 0.438 | 0.909 | 0.467 | 0.419 |
| **\ PART** | 0.394 | 0.031 | 0.524 | 0.923 | 0.450 | 0.414 |
| **\ C4.5 (Only Bayes Net)** | 0.538 | 0.178 | 0.208 | 0.800 | 0.300 | 0.241 |

*Classifiers tested in iterative backward ablation analysis. The first classifier in the first row is the MDT that employs all of the 8 base classifiers. The remaining classifiers are listed in the order in which they were selected to be removed iteratively from MDT for ablation study. '\' indicates "removed." For example, the second classifier is the MDT after SVM was removed, and the third classifier is the MDT after SVM and Voted Perceptron were removed from MDT. The MDT in the final row applied only Bayes Net after C4.5 was removed.

**Table 10.** Results of forward ablation analysis

| MDT* | TPR | FPR | PPV | ACC | F-score | MCC |
|---|---|---|---|---|---|---|
| **+SVM** | 0.368 | 0.006 | 0.847 | 0.944 | 0.513 | 0.536 |
| **+ Voted Perceptron** | 0.417 | 0.008 | 0.824 | 0.946 | 0.553 | 0.563 |
| **+ Random Forest** | 0.455 | 0.011 | 0.789 | 0.947 | 0.577 | 0.575 |
| **+ KNN** | 0.455 | 0.010 | 0.790 | 0.947 | 0.577 | 0.575 |
| **+ C4.5** | 0.466 | 0.011 | 0.789 | 0.948 | 0.586 | 0.582 |
| **+ PART** | 0.464 | 0.011 | 0.779 | 0.947 | 0.581 | 0.576 |
| **+ BayesNet** | 0.469 | 0.012 | 0.779 | 0.947 | 0.586 | 0.580 |
| **+ JRip (F19_BaggingMDT_plus3Dsphere)** | 0.463 | 0.011 | 0.791 | 0.947 | 0.584 | 0.581 |

*Classifiers tested in iterative forward ablation analysis. The first classifier in the first row is the MDT that employs only SVM. The remaining classifiers are listed in the order in which they were selected to be added iteratively to MDT for ablation study. '+' indicates "added." For example, the second classifier is the MDT after SVM and Voted Perceptron were added. The MDT in the final row applied all 8 base classifiers.

on meta-learning, which exploits the synergy among various prediction tools, demonstrated prediction performance superior to that of single epitope predictors. Nevertheless, the current meta-learning approach to epitope prediction depends heavily upon the predictive strength of other pretrained conformational and linear epitope predictors that users cannot retrain directly. This limits the applicability and flexibility of meta-classifiers. Here, we proposed a cost-sensitive bagging MDT approach, which combines two ensemble learning techniques with a cost-sensitive method. This method does not employ the predictions of any pretrained single epitope predictor, making it independent of multiple epitope prediction tools and capable of learning a meta-classification architecture from different given data, rather than restricting it through a prespecified and fixed hierarchy. This method applies the bagging mechanism to reduce the variance in the results of MDTs and considers the misclassification cost to adjust the final prediction and address the imbalanced class distribution in B-cell epitopes.

The structural characteristics of epitopes are different between bound and unbound states. Unlike previous studies of epitope prediction that conducted experiments mainly on bound-state structures [6, 7, 16, 34], we evaluated the performance of MDT for bound and unbound epitope prediction, respectively. While most of the epitope predictors have been trained from and tested on different bound-state antigens, to draw a fair comparison between MDT and other predictors for bound-state antigen prediction, instead of performing k-fold CV, we ran an independent test and also made individual comparisons. The results of the independent test and individual comparisons both demonstrated that our proposed meta-learning MDT approach outperformed the single base tools and other recently developed meta-learning epitope predictors for bound-state epitope prediction. By contrast, we conducted antigen-based 10-fold CV for unbound epitope prediction evaluation between MDT and other epitope predictors. The 10-fold CV was conducted on a recently constructed unbound structure data set [39]. It had not been used yet to pretrain any of the epitope predictors in comparison, and consequently, it was adopted to compare the performance for unbound epitope prediction. The results demonstrated the superior performance of MDT in comparison with three commonly used structure-based epitope predictors, and showed a marginally but noticeably higher performance than that of an unbound-state epitope predictor.

## Acknowledgement

This work was partially supported by Ministry of Science and Technology (MOST) of Taiwan, MOST 106-2221-E-009-184.

## References

1. Meloen RH, Puijk WC, Langeveld JP, Langedijk JP, Timmerman P. Design of synthetic peptides for diagnostics. Curr Protein Pept Sci. 2003; 4: 253-260.
2. Tanabe S. Epitope peptides and immunotherapy. Curr Protein Pept Sci. 2007; 8: 109-118.
3. Naz RK, Dabir P. Peptide vaccines against cancer, infectious diseases, and conception. Front Biosci. 2007; 12: 1833-1844.
4. Benjamin DC, Berzofsky JA, East IJ, Gurd FR, Hannum C, Leach SJ, et al. The antigenic structure of proteins: a reappraisal. Annu Rev Immunol. 1984; 2: 67-101.
5. Pellequer JL, Westhof E, Van Regenmortel MH. Predicting location of continuous epitopes in proteins from their primary structures. Methods Enzymol. 1991; 203: 176-201.
6. Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 2006; 15: 2558-2567.
7. Qi T, Qiu T, Zhang Q, Tang K, Fan Y, Qiu J, et al. SEPPA 2.0-more refined server to predict spatial epitope considering species of immune host and subcellular localization of protein antigen. Nucleic Acids Res. 2014; 42: W59-63.
8. Karplus PA, Schulz GE. Prediction of Chain Flexibility in Proteins – a Tool for the Selection of Peptide Antigens. Naturwissenschaften. 1985; 72: 212-213.
9. Rubinstein ND, Mayrose I, Martz E, Pupko T. Epitopia: a web-server for predicting B-cell epitopes. BMC Bioinformatics. 2009; 10: 287.
10. Zhang W, Liu J, Zhao M, Li Q. Predicting linear B-cell epitopes by using sequence-derived structural and physicochemical features. Int J Data Min Bioinform. 2012; 6: 557-569.

11. Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. BMC Bioinformatics. 2010; 11: 381.

12. Zhang W, Niu Y, Xiong Y, Zhao M, Yu R, Liu J. Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning. PLoS One. 2012; 7: e43575.

13. Hu Y-J, Lin SC, Lin YL, Lin KH, You SN. A meta-learning approach for B-cell conformational epitope prediction. BMC Bioinformatics. 2014; 15: 378.

14. Todorovski L, Dzeroski S. Combining Multiple Models with Meta Decision Trees. Lecture Notes in Computer Science. 2002; 1910: 54-64.

15. Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann Publishers. 2nd edition. 2005. 560.

16. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. PLoS Comput Biol. 2012; 8: e1002829.

17. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. 1985; 55: 836-839.

18. Pellequer JL, Westhof E, Van Regenmortel MH. Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunol Lett. 1993; 36: 83-99.

19. Janin J, Wodak S, Levitt M, Maigret B. Conformation of amino acid side-chains in proteins. J Mol Biol. 1978; 125: 357-386.

20. Ponnuswamy PK, Prabhakaran M, Manavalan P. Hydrophobic Packing and Spatial Arrangement of Amino-Acid-Residues in Globular-Proteins. Biochimica Et Biophysica Acta. 1980; 623: 301-316.

21. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974; 185: 862-864.

22. Ansari HR, Raghava GP. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. Immunome Res. 2010; 6: 6.

23. Breiman L. Random forests. Machine Learning. 2001; 45: 5-32.

24. Chang CC, Lin CJ. LIBSVM: A Library for Support Vector Machines. ACM Trans Intell Syst Technol. 2011; 2: 27.

25. Quinlan JR. C4. 5: Programs for Machine Learning. San Francisco: Morgan Kaufmann Publishers; 1993.

26. Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans Inf Theory. 1967; 13: 21-27.

27. Frank E, Witten IH. Generating Accurate Rule Sets Without Global Optimization. Proceedings of the Fifteenth International Conference on Machine Learning. 1998; 144-151.

28. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco: Morgan Kaufmann Publishers; 1988.

29. Cohen WW. Fast Effective Rule Induction. Proceedings of the Fifteenth International Conference on Machine Learning. 1995; 115-123.

30. Freund Y, Schapire RF. Large margin classification using the perceptron algorithm. Machine Learning. 1999; 37: 277-296.

31. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC press; 1984.

32. Breiman L. Bagging predictors. Machine Learning. 1996; 24: 123-140.

33. Zhang W, Xiong Y, Zhao M, Zou H, Ye XE, Liu J. Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. BMC Bioinformatics. 2011; 12: 341.

34. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008; 9: 514.

35. Zhao L, Wong L, Li J. Antibody-specified B-cell epitope prediction in line with the principle of context-awareness. IEEE/ACM Trans Comput Biol Bioinform. 2011; 8: 1483-1494.

36. Schlessinger A, Ofran Y, Yachdav G, Rost B. Epitome: database of structure-inferred antigenic epitopes. Nucleic Acids Res. 2006; 34: D777-D780.

37. Ponomarenko, J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, Bourne PE. IEDB-3D: structural data within the immune epitope database. Nucleic Acids Res. 2011; 39: D1164-D1170.

38. Ren J, Liu Q, Ellis J, Li J. Tertiary structure-based prediction of conformational B-cell epitopes through B factors. Bioinformatics. 2014; 30: 264-273.

39. Ren J, Liu Q, Ellis J, Li J. Positive-unlabeled learning for the prediction of conformational B-cell epitopes. BMC Bioinformatics. 2015; 16: S12.

40. Hubert L, Arabie P. Comparing partitions. J Classif. 1985; 2: 193-218.

41. Milligan GW, Cooper MC. A study of the comparability of external criteria for hierarchical cluster analysis. Multivariate Behavioral Research. 1986; 21: 441-458.

# Designing Clinical Concept Models for a Nationwide Electronic Health Records System For Japan

**Shinji Kobayashi[1], Naoto Kume[1], Takahiro Nakahara[2] and Hiroyuki Yoshihara[1]**

[1]Department of Electronic Health Record, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2] Kyushu dental University, Fukuoka, Japan

## Abstract

**Objectives:** We developed an electronic health records (EHR) system for regional healthcare in 2000. This EHR stores the health records of more than 6,300 patients in two regions of Japan; however, clinical updates and improved interoperability with other clinical standards, such as HL7 or others are needed. In 2015, this EHR system was upgraded to create a nationwide-scale healthcare data repository to improve the interoperability of clinical data with openEHR technology.

**Methods:** The clinical data in our EHR system has 16 components constructed with Medical Markup Language (MML) standards and periodic mass screening for employees and students. Therefore, we constructed mindmaps of the clinical MML and surveillance data to analyse the concept models. Based on mindmap analysis, we designed archetypes of the concepts identified using Ocean Archetype Editor. The artefacts were mainly quoted from the openEHR clinical knowledge manager (CKM). As the archetypes on CKM did not include all MML semantics, the archetypes were newly designed to complement the semantics of the MML

**Results:** We developed clinical information models by archetypes that semantically equalled the EHR system. Twenty-one MML components/modules and concept models using 99 archetypes were constructed for periodic mass screening services. Most of the archetypes were quoted from CKM; however, 22 archetypes were specialised, and eight archetypes were newly designed. The reasons for specialisation were to adjust the demographics to Japanese and to extend the archetypes to the dental domain.

**Conclusion:** We constructed concept models with archetypes semantically equivalent to conventional data and developed new archetypes for mass screening by archetype technology. The suggested archetype technology improved the flexibility of the EHR system to cover the existing standards.

### Keywords

Archetype; Data modelling; Electronic health records and systems; HL7; Standards adoption; Semantic interoperability

### Correspondence to:

**Dr. Shinji Kobayashi**
Department of Electronic Health Record, Graduate School of Medicine, Kyoto University, Kyoto, Japan.
E-mail: shnj.kobayashi@gmail.com

**EJBI 2018; 14(1):16-21**
Received: November 28, 2017
Accepted: December 30, 2017
published: January 10, 2018

## 1 Background and Significance

In 2003, we developed an electronic health records (EHR) system called the "Dolphin project" for regional health providers to share clinical records and provide individual patient data to patients [1]. We also developed an XML-based clinical information standard called Medical Markup Language (MML) to share clinical data within and among hospitals [2, 3].

We reported that this EHR system was working successfully, with more than 2,100 patients in two regions in Japan in 2010 [4], and 6,300 patients were registered in this system in 2017. Consequently, we launched a project to expand this EHR nationwide to collect clinical data from more hospitals, implement mass screening services for the nation, and reuse the data for research and public health purposes. However, the Dolphin system has three problems that limit its ability to meet the requirements of the next EHR system. First, the Dolphin system has not been sufficiently maintained or updated because it was built on a legacy web framework and was written in Perl language. Second, the Dolphin system adopted MML for the data models, but MML does not meet the clinical/technical demands for emerging clinical updates. Therefore, the Dolphin system lacks interoperability with other standards, such as Health

Level Seven (HL7) [5] and others [6, 7, 8]. Interoperability is an emerging issue for the new project to utilise HL7 standardised data available from healthcare systems.

OpenEHR is the basis of the ISO 13606 [9, 10, 11, 12, 13] standard that enables interoperability of clinical concepts [14] and has been widely adopted to build other standards [15, 16]. The Clinical Information Modeling Initiative, an HL7 working group, is developing a common clinical concept model using openEHR archetype technology [17].

## 2 Objectives

We decided to reconstruct the EHR system using a modern development process, with common clinical concept models based on the ISO 13606 standard with the openEHR specification, to improve flexibility and interoperability for collecting multiple types of formatted data with various standards and catch up requirements of the nationwide EHR project in Japan.

## 3 Method

The data sources for this EHR project were hospitals that were already involved in the Dolphin project and performed periodic mass screening of employees and students. Healthcare data were collected from healthcare providers and screening services. The data were to be formatted using 16 MML content modules and others. MML data modules were analysed and constructed with semantically equivalent archetypes [18]. The mass screening service for employees adopted the Health Level Seven Clinical Document Architecture Release 2 (HL7 CDA R2), so that the surveillance data were collected by HL7 CDA R2 [19]. Because the mass screening service for students did not adopt any standards, raw data were used. Figure 1 shows all datasets analysed with regard to item, structure, components, and concepts. Ultimately 40 mindmaps were constructed.

Based on mindmap analysis, archetypes and templates were designed using Ocean Archetype Editor [20] and Ocean Template Designer [21]. The artefacts were mainly quoted from the openEHR clinical knowledge manager (CKM) [22] with or without modifications to maintain interoperability with regard to other standards. As the archetypes on CKM could not cover all case semantics, the archetypes were newly designed to complement the semantics used.

## 4 Results

All datasets were constructed on 18 templates and mapped with each corresponding data source (Table 1). Archetypes were reused for multiple templates, and Table 2 shows common archetypes in representative cases.

Figure 2 displays an example of the mindmap for a dental examination of teeth in the mass screening for students.

Table 1 shows each case and the number of archetypes used to construct semantically equivalent data schema. All components were designed with 1–50 archetypes and were fully conformant with openEHR specifications [23]. The most complex model was the mass screening for students, which required 50 archetypes. Mass screening for students requires a dental examination (Table 1). Figure 3 provides an overview of the archetype hierarchy in for mass screening of students. The report information model was designed for use with general reporting documents, but it had to be redesigned for use with specific reports, such as imaging reports, pathological diagnosis reports, and other diagnostic reports.

Table 3 shows the details of archetypes used to construct logical data models. A total of 92 archetypes were needed to construct the models, and 62 archetypes were quoted from CKM without any changes. However, 22 archetypes had to be

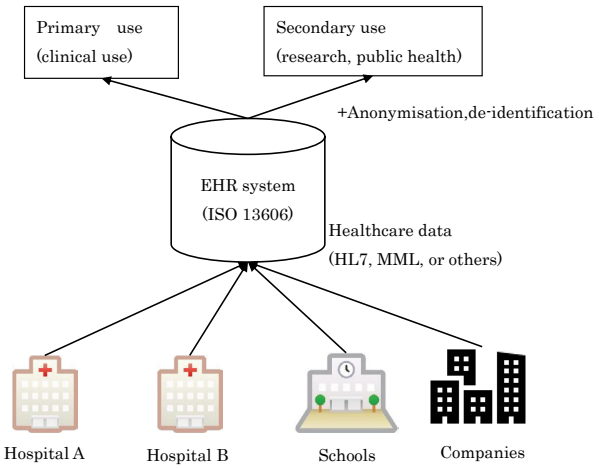



Figure 1: Nationwide electronic health records (EHR) system. Healthcare data will be collected from hospitals, schools, and companies by HL7, MML, original format for mass screening, or others. The EHR system stores healthcare data in the ISO 13606/openEHR repository for primary/secondary use with/without anonymisation or de-identification.

Table 1: Number of archetypes used for each concept.

| Use case | Number of archetypes |
|---|---|
| Personal information | 8 |
| Creator information | 7 |
| Patient information | 9 |
| Health insurance | 10 |
| Diagnosis record | 2 |
| Lifestyle information | 5 |
| Basic clinical information | 5 |
| Initial consultation | 14 |
| Progress course | 15 |
| Surgical operation record | 13 |
| Clinical summary | 19 |
| Laboratory test | 11 |
| Report information | 18 |
| Referral letter | 22 |
| Mass screening for employee | 36 |
| Mass screening for students | 50 |

Each case was analysed and constructed as an archetype to a template.

Table 2: Frequently reused archetypes in multiple cases.

| Archetypes | Total | Diagnostic report | Progress notes | Clinical summary | Surgical operation record | Flow sheet | Referral | Mass screening for employee | Mass screening for students |
|---|---|---|---|---|---|---|---|---|---|
| CLUSTER.organisation-mml | 26 | 6 | | 6 | 5 | 3 | 6 | | |
| CLUSTER.citation | 21 | 1 | 8 | 7 | 1 | | 4 | | |
| EVALUATION.problem_diagnosis | 15 | 1 | 1 | | 1 | | 1 | 2 | 9 |
| EVALUATION.citation | 13 | 1 | 4 | 5 | | | 3 | | |
| CLUSTER.person_name-mml | 12 | 3 | | 3 | 4 | | 2 | | |
| CLUSTER.individual_professional-mml | 10 | 3 | | 3 | 2 | | 2 | | |
| EVALUATION.clinical_synopsis | 10 | 1 | 1 | 3 | 1 | 2 | 2 | | |
| CLUSTER.device | 9 | | | | | 4 | | | 5 |
| EVALUATION.health_risk | 8 | | | | | | | 8 | |
| OBSERVATION.progress_note | 8 | | 3 | 2 | | | 3 | | |
| INSTRUCTION.medication_order | 7 | | 4 | 1 | | | 2 | | |
| CLUSTER.address-japan | 7 | | | 3 | 2 | | 2 | | |
| OBSERVATION.exam | 6 | | 1 | 1 | | | | 2 | 2 |
| CLUSTER.telecom_details-mml | 6 | | | 2 | 2 | | 2 | | |
| OBSERVATION.story | 5 | | 1 | 1 | | | 2 | 1 | |
| OBSERVATION.body_weight | 4 | | | | | 1 | | 2 | 1 |
| INSTRUCTION.request | 4 | 1 | 3 | | | | | | |
| CLUSTER.medication_preparation | 4 | | 2 | 1 | | | 1 | | |
| CLUSTER.daily_timing | 4 | | 2 | 1 | | | 1 | | |

The common archetypes listed above were used for multiple cases. Organisation archetypes are commonly used to describe the facility, department, or ward. Citation archetypes are used to point to external links to images or other relative resources. Problem/diagnosis archetypes are also frequently reused.

modified, and eight archetypes were created from scratch to conform to the concept models. Reasons for modifying archetypes had to do mainly with localising for demographics, insurance claims, and the inclusion of specific information for dentistry (Table 4).

Because Health Level Seven version 2.5 (HL7 2.5) messages were also mapped to the constructed archetypes, this EHR system obtained the HL7 2.5 interface. Although the interface of the old EHR system was only MML, the newly developed EHR system used the following interfaces mapped to ISO 13606/openEHR archetypes:

- ISO 13606/openEHR native API
- MML

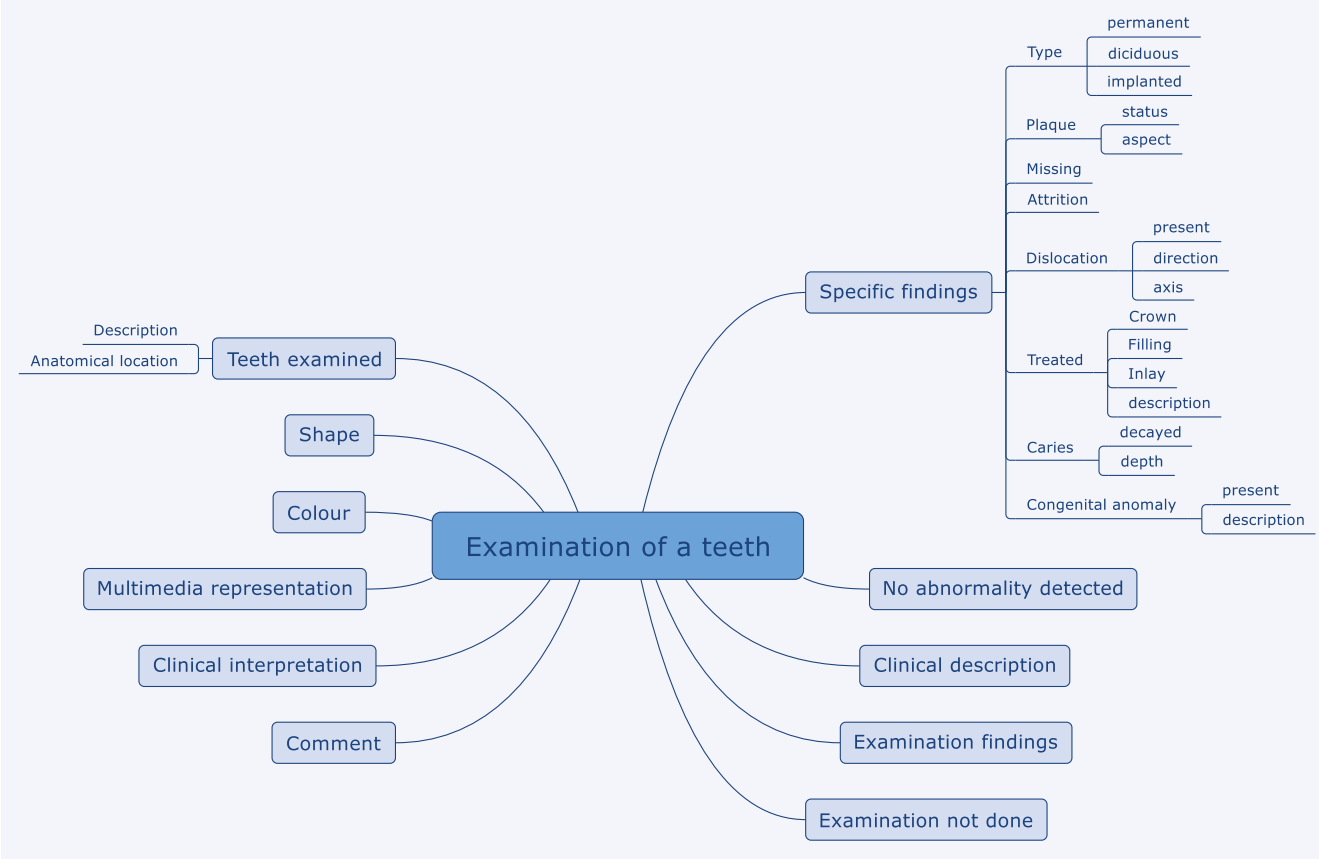



Figure 2: Mindmap for the examination of teeth.

A concept model for clinically examining teeth was developed into a mindmap. The concept was taken from the openEHR Clinical Knowledge Manager and specialised for dental examination-specific items.

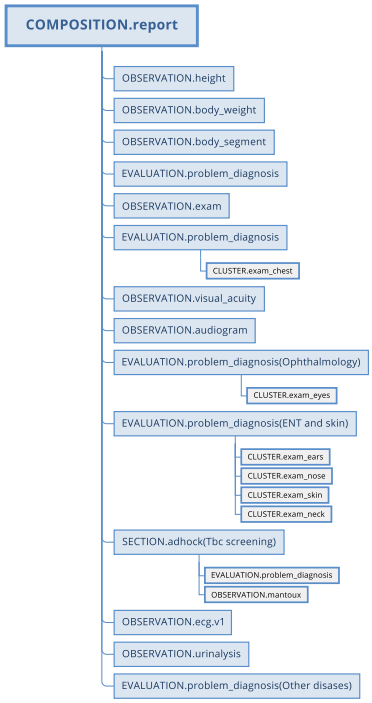



Figure 3: Overview of the hierarchy of archetypes in mass screening for students.

Table 3: Archetype construction using the case information model.

| Archetype | Count of use |
|---|---|
| No change | 62 |
| Specialised | 22 |
| Newly created | 8 |

Archetypes were mainly derived from the openEHR clinical knowledge manager. Sixty-two archetypes were used as-is, ten required specialisation, and eight archetypes were newly created.

Table 4: Example of rationales for specialised archetypes.

| Concept model | Rationale |
|---|---|
| Patient name | To record ideographic (Kanji), syllabic (Furigana), and alphabetic names |
| Address | To record addresses in ideographic, syllabic, and alphabetic form<br>To add the field for prefecture code |
| Telecoms | To adjust the item hierarchy to suit Japanese telecoms |
| Procedure | To match claim information to insurance systems |
| Medication | To record Japanese-style preparation and claims information for insurance purposes |
| Report | To record the report date, name of requester, responder to context field |
| Dental examination | To record specific expression for dentists |

- HL7 2.5
- HL7 CDA R2
- Original data format for mass screening

## 5 Discussion

An EHR system requires semantic interoperability for multiple use cases, such as point of care, clinical research, and public health [24]. To achieve interoperability, EHR should have a commonly agreed upon logical model [25]. Therefore, the openEHR project has developed the CKM to share archetype concept models [26] to improve the interoperability of EHR.

The CKM is also a repository for international collaborative development of clinical concepts, and the CKM reported that 679 archetypes have been developed (cited on 19 December 2017). Archetypes on CKM are designed for general international use; however, in our case, the archetypes had to be localised mainly to express local demographics as well as the inclusion of dental examinations. As demographic expressions in CKM are widely diverse from MML, it was not possible to cover all of them. Specialising the existing archetypes was a reasonable way to localise them. Thus, although the existing HL7 and MML data models captured in the archetypes had an impedance mismatch in terms of information granularity, it was possible to devise equivalent forms with 92 archetypes; in fact, 67% (62/92) of them were adopted with no modifications. Although 1,821 users are registered with CKM, there are only 34 dentists (CKM statistics page, retrieved on 19 December 2017). Thus, there are few archetypes for dentistry information on CKM, and we had to modify the archetypes to record data from dental examinations.

Moreover, the artefacts can be modified with other archetypes to describe other cases immediately, suggesting that creating a standardised clinical concept in the openEHR project would have more potential than making an EHR system with existing data models. However, we needed more clinical concepts to meet the unique demand of local clinicians and specialists.

Because archetype concept models have fine-grained items, they are capable of covering multiple standardised formats, such as HL7 2.5 and HL7 CDA, among others. As archetypes must be interoperable, the concept models in this study could be used with other standards. In addition, the old EHR system had only an MML interface, but the new EHR system uses five interfaces with archetype technology.

We are currently implementing an EHR system with these archetype concept models and, in the near future, plan to demonstrate interoperability for secondary use.

## 6 Conclusion

We constructed concept models with archetypes for a nationwide EHR system, which is semantically equivalent to the legacy EHR. Mass screening of the newly developed archetypes was conducted, and 11 dental archetypes were modified based on existing archetypes. The suggested archetype technology improved flexibility for collecting health data from multiple existing standards.

## 7 Clinical Relevance Statement

The developed clinical concept models can be used to describe clinical contents. Using these models would clarify the semantics of clinical descriptions and assessments and improve machine processing for clinical data. Thus, this project should contribute to clinical data management and the aid in the discovery of novel findings from EHRs.

## 8 Conflict of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## 9 Human Subjects Protections

The study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and Japanese laws and guidelines for privacy and security.

## 10 Acknowledgements

This research was supported by the "Clinical Study Oriented ICT Infrastructure Development Project -

Sustainable Massive Health and Clinical Data Repository for Secondary Use" of the Japan Agency for Medical Research and Development, AMED.

## References

[1] Takada A, Guo SJ, Tanaka K, Sato J, Suzuki M, Suenaga T. et al. Dolphin project--cooperative regional clinical system centered on clinical information center. J Med Syst. 2005; 29: 391-400.

[2] Guo J, Araki K, Tanaka K, Sato J, Suzuki M, Takada A. et al. The latest MML (Medical Markup Language) Version 2.3 - XML-based standard for medical data exchange/storage. J Med Syst 2003; 27: 357–66.

[3] Guo J, Takada A, Tanaka K, Sato J, Suzuki M, Suzuki T. et al. The development of MML (Medical Markup Language) version 3.0 as a medical document exchange format for HL7 messages. J Med Syst 2004; 28: 523-33.

[4] Takemura T, Araki K, Arita K, Suzuki T, Okamoto K, Kume N. et al. Development of fundamental infrastructure for nationwide EHR in Japan. J Med Syst. 2012; 36: 2213–8.

[5] Health Level Seven® International. http://www.hl7.org/

[6] Integrating the Healthcare Enterprise (IHE). https://www.ihe.net/

[7] Clinical Data Interchange Standards Consortium. https://www.cdisc.org/

[8] DICOM® (Digital Imaging and Communications in Medicine). http://www.dicomstandard.org/

[9] ISO 13606-1:2008. Health informatics -- Electronic health record communication -- Part 1: Reference model.

[10] ISO 13606-2 :2008. Health Informatics – Electronic health record communication – Part 2: Archetype interchange specification.

[11] ISO 13606-3:2009. Health informatics -- Electronic health record communication -- Part 3: Reference archetypes and term lists.

[12] ISO/TS 13606-4:2009. Health informatics -- Electronic health record communication -- Part 4: Security.

[13] ISO 13606-5:2010. Health informatics -- Electronic health record communication -- Part 5: Interface specification.

[14] Beale T. Archetypes: Constraint-based domain models for future-proof information systems. Proceedings of Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer. 2002: 16-32.

[15] ISO 13940:2015. Health informatics -- System of concepts to support continuity of care.

[16] Goossen WTF. Detailed clinical models: Representing knowledge, data and semantics in healthcare information technology. Healthc Inform Res. 2014; 20: 163–72.

[17] Policies: Selected CIMI policies and decisions. http://opencimi.org/policies_cimi

[18] Kobayashi S, Bosca D, Kume N, Yoshihara H. Reforming MML (Medical Markup Language) standard with archetype technology. Indian J Med Inform. 2014; 8: 57–60.

[19] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron P V. et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform. 2006; 13: 30–9.

[20] Informatics O. Ocean Archetype Editor. http://wiki.oceaninformatics.com/confluence/display/TTL/Archetype+Editor+Releases

[21] Informatics O. Ocean Template Designer. http://www.openehr.org/downloads/modellingtools

[22] openEHR Clinical Knowledge Manager. http://www.openehr.org/ckm/

[23] The openEHR. openEHR specification program; 2015. http://www.openehr.org/programs/specification/latestreleases

[24] ISO/TS 18308: 2011. Health informatics — Requirements for an electronic health record architecture.

[25] ISO/TR 20514: 2005. Health informatics —Electronic health record —Definition, scope and context.

[26] Garde S, Chen R, Leslie H, Beale T, McNicoll I, Heard S. Archetype-based knowledge management for semantic interoperability of electronic health records. Stud Health Technol Inform. 2009; 150: 1007–11.

# Zika Virus Viewed Through the Resonant Recognition Model. Unraveling New Avenues for Understanding and Managing a Serious Threat

**José Luis Hernández Cáceres[1*] and Graham Wright[2]**

[1] Cuban Neuroscience Center, Havana, Cuba

[2]Graham Wright, Department of Information Systems, Rhodes University, Grahamstown, South Africa

## Abstract

**Introduction**: Zika Virus (ZIKV) infection is a major public health concern, affecting almost each country in the western hemisphere. A more than 20-fold increase in microcephaly risks is associated to ZIKV infection in pregnancy. A new vaccine is not expected before 2019, and alternative prophylactic and therapeutic approaches are encouraged. We expect that the Resonant Recognition Model, developed by Irena Cosic, might lay on the basis for an alternative approach to handle ZIKV.

**Objective:** We tried to identify the resonant frequencies associated to the ZIKV polyprotein and their use for an automatic taxonomy of different ZIKV strains. We put to test the hypothesis of interaction between ZIKV envelope protein and the AXL receptor, one of the plausible mechanisms proposed for ZIKV-associated microcephaly.

**Results:** Four relevant frequencies ($f_{RRM}$) were found in ZIKV polyprotein consensus spectrum. Corresponding four spectral amplitudes allowed separating African from Asian/American ZIKV isolates (k-means clustering). Peak 3 ($f_{RRM}$= 0.2754) and Peak 4 ($f_{RRM}$= 0.334) yielded a finer separation between Asian sequences and those from the current outbreak collected in 2015 (Asian/American). Consensus spectrum for pooled Dengue virus and ZIKV polyprotein sequences suggest that Peak 4 might be a specific hallmark of ZIKV. RRM results support the interaction between ZIKV envelope protein and AXL membrane receptor. The interacting frequency of $f_{RRM}$= 0.167 seems to be a sub-harmonic of Peak 4.

**Conclusions:** Resonant recognition model provides a plausible view of processes involved in the interactions of ZIKV with the human host, and is suggesting the exchange of electromagnetic radiation at the frequencies of 601.8nm (yellow light) and 1203.6 (near infrared) during ZIKV envelope protein with the AXL receptor in the human fetal tissue. This information might be relevant for alternative approaches to new therapeutic approaches to treat ZIKV-associated damage to newborns.

### Keywords

Zika virus; AXL; Resonant recognition model; Flaviviridae

### Correspondence to:

**José Luis Hernández Cáceres,**
PhD, Cuban Neuroscience Center, Havana, Cuba.
E-mail: Kacerjlh@gmail.com

**EJBI 2018; 14(1):11-17**
Received: November 06, 2017
Accepted: December 26, 2017
Published: January 10, 2018

## 1 Introduction

In 1947, Rhesus 766, a captive monkey in the Zika Forest of Uganda was detectedwith fever.Two days later, Rhesus 766, still febrile, was brought to the Rockefeller Foundation's laboratory at Entebbe and its serum wasintra-cerebrally inoculated into mice. After 10 days, all mice were sick, and a filterable transmissible agent, later named Zika virus (ZIKV), was isolated fromthe murine brains. In early 1948, ZIKV was also isolated from *Aedes africanus* mosquitoes captured in the Zika forest. In 1952, for first time, a case of human infection with Zika virus was documented in Uganda [1]. In 1964, Simpson provided the first well-documented report of human ZIKV disease (ZVD), narrating his own occupationally acquired Zika fever (ZF). Disease beganwith minor headache, followed the next day by maculopapular rash covering face, neck, trunk, and upper arms, and spreadingto palms and soles. The clinical picture also included fever, malaise, and back pain. By the evening of the second dayfever disappeared, the rash was receding, and by day three, he felt well. Rash completely disappeared over the next two days [2].

Until 2007, little attention was drawn to the ZVD. Its course was mild, and occurred mainly in a geographical region where malaria, HIV, tuberculosis, malnutrition and measles were the major reasons forinternational public health concern. According to some views, forest monkeys are the natural reservoir for the virus, which is transmitted among individuals by forest *Aedes* mosquitoes, with humans only as incidental hosts. Anthropological intervention apparently altered the sylvatic cycle of transmission: the virus mutated and started a dangerous spread around the planet. By 2007 it reached the Yap Island, in the Federate States of Micronesia, where 75% of their inhabitants fell infected.

At that moment (just tenyears ago!) premonitory were the words of Edward B. Hayes: "Fortunately, ZIKV illness to date has been mild and self-limited, but before West Nile virus caused large outbreaks of neuroinvasive disease in Romania and in North America, it was also considered to be a relatively innocuous pathogen" [3, 4].

After spreading to Malaysia and Cambodia, ZIKV finally reached South America by 2014, and by 2015 completed the circumnavigation of the planet reaching Cape Verde Islands in Africa.

Even when ZF still remains as a mild or even clinically silent entity, its most alarming feature is the proven association with certain neurological disorders, especially Guillain Barre Syndrome in adults and microcephaly in babies born to mothers infected during pregnancy (Figure 1).

From February 1st, 2016 until October 30th 2016, Zika virus was regarded by World Health Organization as a Public Health Emergency of International Concern (PHEIC) [5]. The figures of ZVDcases reached up to 6 digits globally [6, 7, 8], and the hypothesis of association of ZIKV with microcephaly as well as Guillain-Barre and ophthalmic anomalies is receiving mounting support [9]. In the Americas, all countries except Canada have been affected; in continental USA 5,575cases were reported up to September 30, 2017 [10].

ZVD is of particular concern among pregnant women. Recent reports suggest that Zika infection might increase the risk of brain malformation by a factor of 20-70 [11, 12].

A vaccine for ZIKV is not expected before 2019, and currently available antivirals are ineffective. There are worries regarding the efficacy of a ZIKV vaccine for curbing a ZVD pandemic. This justifies the search for alternative ways for ZIKV combat. As stated by Malone et al. [13]: "In the absence of currently available vaccines, the likely long timeline for vaccine development, and the open questions about the basic pathogenesis of Zika virus infection, parallel development of other prophylactics and therapeutics must be explored."

In developing countries, Bioinformatics is looming as one of the most promising and cost-effective research strategies. The application of mathematical models, the development of data mining and other similar approaches to the study of the huge genomic databases of public domain can devise new strategies aimed at understanding as well as proposing new ways of combating diseases.

Here we are showing how an approach based on the resonant recognition model can contribute to new proposals for facing ZVD.

The application of ideas and concepts from the Resonant Recognition Model (RRM) for infectious diseases control has been proposed for treating Ebola Virus Disease [14] as well as for malaria [15, 16]. Experimental evidence supporting the resonant exchange of electromagnetic energy as a way of interaction between proteins and their receptors, substrates and regulators (as postulated by RRM) is ample [17, 18 19], ideas based on this approach might propose new therapeutic ways to treat ZVD.

The existence of a mouse model for ZF [20], as well as the possibility to apply affordable low power radiation sources (e.g. LED´s [14]) might provide support to new strategies based on RRM. Drug design based on RRM has also suggested as an alternative way for treating different diseases, as some types of cancer [21].

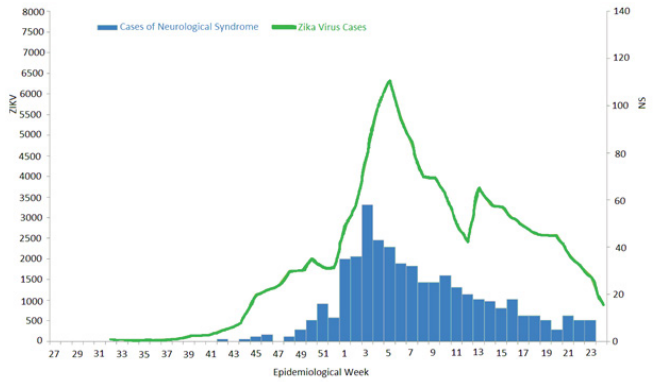



Figure 1: Cases of Neurological Syndrome and suspected and confirmed Zika virus disease cases by epidemiological week as reported for Colombia, 2015-2016. Modified from [4]

The main postulate of the RRM is that proteins do recognize their substrates/and receptors via exchanging electromagnetic radiation [17, 18, 19]. This process follow a resonant mechanism, and the exchanged energies are in the range of $10^{-20}$—$10^{-18}$ J (0.1—10 eV) [22].

A key aspect of the RRM approach is to identify the resonant frequencies involved. For it, one or various proteins that are determinative for fulfilling a given function were analyzed. As result of RRM analysis relevant frequencies as well as putative interactions might be identified. Below, a group of proteins related to ZIKV infection have been submitted to RRM analysis. Our results point to the plausibility of RRM as a way to understand some of the ZIKV mechanisms and, hopefully, to devise new therapeutic strategies.

# 2 Methods

## 2.1 Sequences

**Flaviviridae polyproteins**

Zika and other arboviruses´ genomes contain a coding region for a polypeptide which is processed into structural capsid, premembrane/membrane, envelope and non-structural proteins [6, 7, 8]. The following ZIKV full polyprotein sequences were downloaded from GenBank database (as translated into amino acid sequences from original full genome RNA sequences; Table 1).

They corresponded to the full list of ZIKV polyprotein sequences available at GenBank up to February 2016 [23].

The following polyprotein sequences from dengue virus were also downloaded:

P27909 (POLG_DEN1B)

P17763 (POLG_DEN1W)

P27912 (POLG_DEN1A)

P14337 (POLG_DEN28)

P29991 (POLG_DEN27)

## 2.2 Envelope Protiens

ZIKV envelope protein plays an important role in viral infection. It has been associated to functions such as fusion of virus membrane with host endosomal membrane, host-virus interaction, viral attachment to host cell, viral penetration into host cytoplasm, and virus entry into host cell [8].

The following six ZIKV envelope sequences were studied:

W8QFD5 – W8QFD5_ZIKV

W8R1N8 – W8R1N8_ZIKV

W8Q6P9 – W8Q6P9_ZIKV

W8Q6P1 – W8Q6P1_ZIKV

W8QFC5 – W8QFC5_ZIKV

W8QIQ1 – W8QIQ1_ZIKV

Each sequence contained 251 AA. Mutations were found in 12 different positions (8%).

Table 1: ZIKV polyprotein sequences analyzed in present study.

| Accession Number | Country | Year of Collection |
|---|---|---|
| LC002520 | Uganda | 1947 |
| HQ23449 | Malaysia | 1966 |
| HQ234500 | Nigeria | 1968 |
| KF383116 | Senegal | 1968 |
| KF383115 | CAR | 1968 |
| HQ234501 | Senegal | 1984 |
| KF268948 | CAR | 1976 |
| KF268949 | CAR | 1980 |
| KF268950 | CAR | 1980 |
| KF383117 | Senegal | 1997 |
| KF383118 | Senegal | 2001 |
| KF383119 | Senegal | 2001 |
| EU545988 | Micronesia | 2007 |
| JN860885 | Cambodia | 2010 |
| KU681082 | Philippines | 2012 |
| KU681081 | Thailand | 2014 |
| KJ776791 | Polynesia | 2015 |
| KU365779 | Brazil | 2015 |
| KU365778 | Brazil | 2015 |
| KU365777 | Brazil | 2015 |
| KU3657780 | Brazil | 2015 |
| KU312312 | Suriname | 2015 |
| KU501215 | P. Rico | 2015 |
| KU509998 | Haiti | 2014 |
| KU321639 | Brazil | 2015 |
| KU527068 | Brazil | 2015 |
| KU647676 | Martinique | 2015 |
| KU501216 | Guatemala | 2015 |
| KU501217 | Guatemala | 2015 |
| KU707826 | Brazil | 2015 |

***Axl***

The tyrosine kinase Axl protein is a cell surface receptor. Experimental evidence suggest that Axl is a strong candidate for ZIKV infection target [24].

The six following Axl sequences were downloaded:

NM_001278599

NM_001699

NM_021913

NM_001190974

NM_001190975

NM_009465

**Resonant recognition model**

The primary amino acid sequence was transformed into a numerical sequence following the Resonant Recognition Model (RRM) methodology [17]. Each of the 20 amino acids in the entire sequence was assigned an electron-ion interaction potential (EIIP) value (Table 2)

The obtained numerical sequence was treated as a time series. Power spectrum was obtained for each sequence using the freely available SciLab cross-platform numerical computational package (https://www.scilab.org/). For finding the consensus spectrum, all the spectral vectors were submitted to scalar cross multiplication. The obtained product is considered as the consensus spectrum. Relevant peak(s) are taken as important frequencies associated to resonant processes associated with the analyzed sequences.

Table 2: EIIP values for each amino acid [17] .

| Amino Acid | EIIP |
|---|---|
| S | 0.0829 |
| T | 0.0946 |
| Q | 0.0761 |
| Y | 0.0516 |
| G | 0.005 |
| A | 0.0373 |
| V | 0.0057 |
| L | 0 |
| I | 0 |
| C | 0.0829 |
| M | 0.0823 |
| P | 0.0198 |
| F | 0.0946 |
| W | 0.0548 |
| K | 0.0371 |
| D | 0.1263 |
| E | 0.0058 |
| R | 0.0959 |
| H | 0.0242 |
| N | 0.0036 |

EIIP represents the average energy state of all of the valence electrons associated with that amino acid.

The RRM frequency ($f_{RRM}$) was converted to a true electromagnetic frequency by determining the appropriate wavelength (λ) using the expression proposed by Cosic [17]:

$$f_{RRM} = 201/\lambda$$

**Statistical processing**

Procedures included cluster, regression and correlation analysis.

## 3 Results

### 3.1 RM Analysis of ZIKV Full Genome Polyproteins

The consensus spectrum corresponding to the 30 known polyprotein sequences available at GenBank in February 2016 allowed to identify four distinct peaks at RRM frequencies ($f_{RRM}$) of 0.038 (Peak 1), 0.125 (Peak 2), 0.2754 (Peak 3), and 0.334 (Peak 4) (Figure 2). These peaks could be regarded as the main candidates for further RRM analysis. These frequencies most likely correspond to wavelengths of 5289.5 nm (mean infrared) for Peak 1; 1608 nm (near infrared) for Peak 2; 729.9 nm (visible red light) for Peak 3, and 601.8 nm (yellow color) for Peak 4.

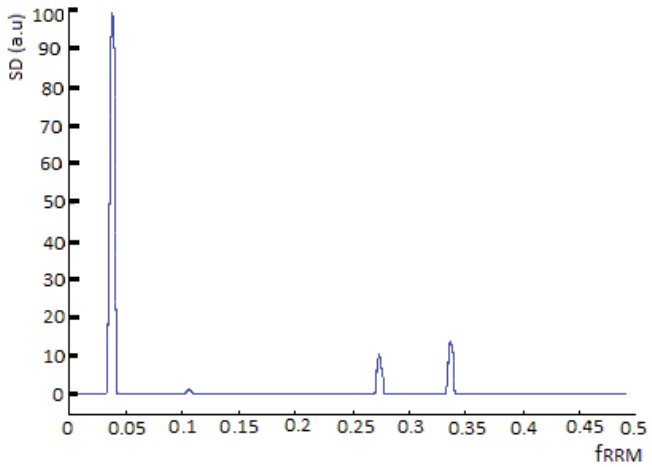



Figure 2: Consensus spectrum corresponding to the 30 ZIKV full genome sequences known up to February2016.Note the four distinct peaks encountered: Peak 1 ($f_{RRM}$ =0.038), Peak 2 ($f_{RRM}$= 0.125), Peak 3 ($f_{RRM}$= 0.2754) and Peak 4 ($f_{RRM}$= 0.334).

## 3.2 Spectral Density of Each Frequency and ZIKV Evolution

According to the RRM, increase or decrease in RRM spectral amplitude is regarded as an enhancement in the related activity/ interaction. Thus we are analyzing in this section possible implications of spectral density values corresponding to a group of 21 ZIKV polyprotein sequences

With the aim of associating spectral density values of each peak to the origin of ZIKV strains, data were submitted to multivariate cluster analysis (k-means clustering), using the amplitude of each spectral peak as the four predictive variables. The following two clusters were obtained:

**Cluster Number 1, containing 10 strains**

- Malaysia (1966)
- Nigeria (1968)
- Senegal (1968)
- Central African Republic (1968)
- Senegal (1984)
- Central African Republic (1976)
- Central African Republic (1980)
- Senegal (1997)
- Senegal (KF383113; 2001)
- Senegal (KF383114;2001)

**Cluster Number 2, containing 11 strains**

- Micronesia (2007)
- Cambodia (2010)
- Polynesia (2015)
- Brazil (KU365771; 2015)
- Brazil (KU365772; 2015)
- Suriname (2015)
- Puerto Rico (2015)
- Haiti (2014)
- Brazil (KU527061; 2015)
- Martinique (2015)
- Guatemala (2015)

This statistical classification based on RRM analysis is in agreement with the view of classifying all African sequences and the Malaysian 1966 isolate as "African" [23], whereas the American sequences of current outbreak are similar to the "Asian" lineage [6, 8].

## 3.3 Relation of Different Peaks to ZIKV Classification

Correlation analysis revealed that Peak 1 and Peak 2 tend to be lower in Asian/American isolates, whereas Peak 3 and Peak 4 are increased among Asian/American ZIKV Polyprotein sequences (Table 3).

## 3.4 Possibility to Discriminate Between Asian and American ZIKV Lineages

Inspection of data from Asian and American ZIKV Polyprotein strains suggested that peaks P3 and P4 seem to increase among the strains from the current epidemic as compared to the Asian outbreaks from2007-2010. It is possible that the virulence of Asian/ American strains is associated to some of these two peaks.

In order to reveal possible numerical discrimination capability, Asian and American ZIKV Polyprotein sequences were submitted to cluster analysis this time using amplitudes of Peak 3 and Peak 4 as predicting variables.

K-means clustering reveled the following two clusters:

**Cluster Number 1, containing 2 cases**

- Micronesia (2007)
- Cambodia (2010)

**Cluster Number 2, containing 8 cases**

- Polynesia (2015)
- Brazil (KU365771; 2015)
- Brazil (KU365772; 2015)
- Suriname (2015)
- Puerto Rico (2015)

Table 3: Significant correlations between ZIKV Polyprotein consensus spectrum peaks and their relation to classification (African corresponds to class 1 whereas Asian/American corresponds to class 2).

| | P2 | P3 | P4 | CLASS |
|---|---|---|---|---|
| P1 | 0.45 | | -0.54 | -0.56 |
| P2 | | -0.63 | -0.44 | -0.93 |
| P3 | | | 0.63 | 0.75 |
| P4 | | | | 0.57 |

- Haiti (2014)
- Brazil (KU527061; 2015)
- Martinique (2015)
- Guatemala (2015)

Plausibly, the Polynesia 2015 isolatewas grouped with other American sequences from the current outbreak, but distinguished from the Cambodian and Micronesian strains from previous outbreaks [23].

### 3.5 Relation to Other Flaviviruses

A consensus spectrum was obtained for a set comprising five dengue virus and 9 ZIKV polyprotein sequences. The consensus spectrum is represented in figure 3. Surprisingly, three of the four peaks found for ZIKV are preserved in this consensus spectrum. Apparently, Peak 4 is negligible, suggesting that this peak might be more specific for ZIKV, whereas Peak 1, Peak 2 and Peak 3 correspond to mechanisms that are common to other flaviviruses.

### 3.6 Exploring Possible Interaction Between Zika Envelope Protein and Tyrosine Kinase Receptor Axl.

Axl has been suggested as a strong candidate receptor to ZIKV envelope protein during infection, in particular during fetal infection leading to microcephaly [24].

One of the keys postulates of the resonant recognition model proposed by Cosic states that interacting molecules „communicate“ with each other i.e. recognize each other at a distance, on the basis of the same characteristic frequency but opposite phases at that frequency. This key aspect of RRM has found experimental confirmation during the study of cross reactivity of certain antibodies: significant cross reactivity to the polyclonal antibodies raised against peptides which share at least one Characteristic frequency and phase at a given frequency has ben be observed [17]. This criterion has been used to determine whether two proteins should interact, as it was shown recently for putatively interacting Plasmodium proteins [16].

Based on the idea that, interacting proteins share a common frequency in the consensus spectrum as well as opposite phases, we tried to test the hypothesis of Axl as a receptor for ZIKV, via RRM analysis of the interaction between ZIKV envelope protein and the Axl protein.

In Figure 4, the consensus spectrum for ZIKV envelope protein and AXL is represented. Peaks appeared at and at $f_{RRM}$=0.072. Curiously, the main peak at $f_{RRM}$= 0.167 seems to appear as a sub-harmonic of Peak 4 (0.334/2=0.167).

A phase shift of 176° (0.978π, very close to the opposite phase condition) was found between ZIKV envelope and AXL, thus supporting an interaction between both proteins (Figure 5). The interacting RRM frequency of 0.167 corresponds to a wavelength of 1203.6 nm (near infrared).

## 4 Discussion

The application of a new, now widely accepted method requires a cautious approach, since prediction of such a method cannot be taken for granted. In our view, first we need to search for the plausibility of obtained data using the innovative RRM approach. Briefly, it can be commented that our previous report on a small sample of ZIKV envelope proteins, detected a major peak at position 0.295 (as evident from Figure 1 of that paper [25]). This frequency was found by Veljkovic et al. [26] to be the main RRM frequency, as they used a larger sample of sequences.

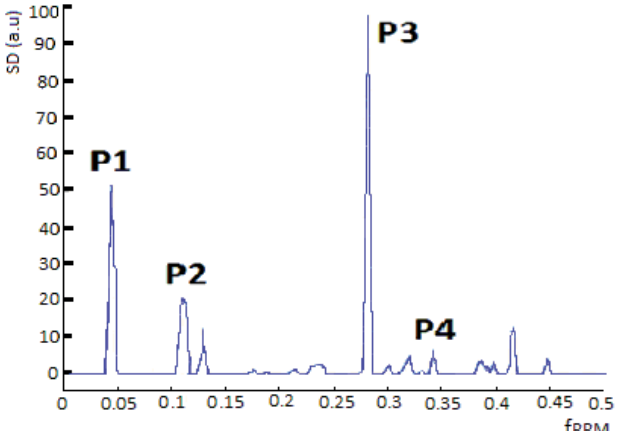



Figure 3: Consensus spectrum for pooled polyprotein sequences from five dengue viruses and nine ZIKV isolates. Peaks 1, 2 and 3 are prominent, whereas Peak 4 is barely present.

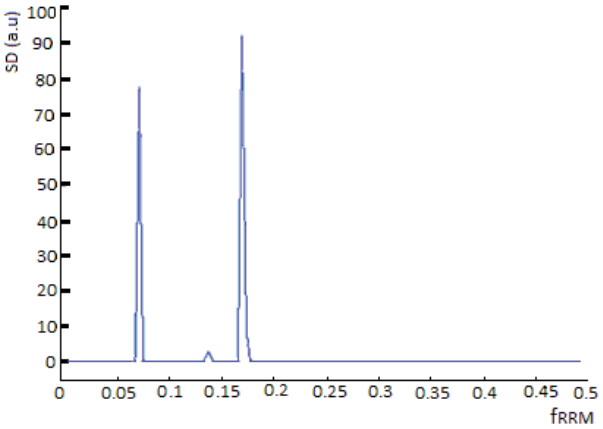



Figure 4: Consensus spectrum for the five ZIKV envelopesequences and six AXL sequences. Note a prominent peak at $f_{RRM}$=0.167and another peak at $f_{RRM}$ =0.072.

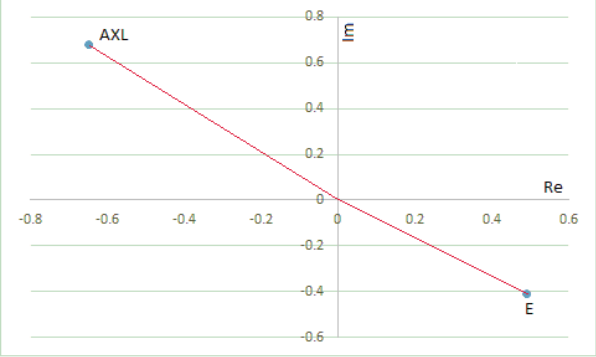



Figure 5: Phase values for ZIKV envelope protein (E) and human Axl receptor protein at $f_{RRM}$= 0.167. A phase shift of 176° (0.978π) was obtained.

In the present study, we are using a standard freely available program for spectrum determination, and the lack of an *ad hoc* package for RRM analysis could make our results incompatible with mainstream research in this area. Fortunately, this does not seem to be the case.

In a recent review, Cosic et al. [27] provided a detailed account of different biological processes associated to different RRM

frequencies. Comparing the frequencies described in this work with those described by Cosic et al. suggests the following:

The peak P1 at 0.038 is closely related to the frequency of 0.039, corresponding to IL-1, a protein involved in cell recognition, cell immunity and inflammation. As it is well known, these processes are common to flavoviruses invasion.

Peak P2 (0.125) is associated to myxoma viral activity (0.115) and is in a RRM frequency region associated to viral and bacterial infection processes [28].

Peak P4 at 0.334 is shared by several plasmodium proteins and apparently is associated to pathogen association to the cell membrane. Notice-worth is that Peak 4 seems to be specific for ZIKV and not to other flaviviruses, and microcephaly has been documented as associated to ZIKV and not to dengue virus.A sub-harmonic of this peak (0.167) has been shown here to correspond to a resonant interaction between ZIKV envelope protein and Axl. The resonant frequency of 0.167 corresponds to the DNA binding/ regulation region. As it has been reported, Axl dose dependently and specifically enhances NF-kB DNA-binding activity [29].

The common peak between AXL and ZIKV envelope protein at 0.072 is closely related to the peak at 0.07 reported for neurotoxins, and might reflect the interaction of ZIKV with neural tissue.

These reasons, added to the fact that these peaks were useful for discriminating between ZIKV strains of different origin do support the plausibility of RRM analysis for understanding molecular mechanisms involved in ZVD.

This report seems to be the first wide-range RRM study of Zika virus polyprotein. Two previous papers related to ZIKV envelope proteins were published recently [25, 26].

The study of 30 polyprotein ZIKV sequences allowed identifying four peaks. The peaks properly allowed to numericallydistinguish between African and Asian/American ZIKV isolates, and this support the idea of a functional role for corresponding electromagnetic frequencies.

Peak 3 and Peak 4, however, seemed to be more parsimonious in discriminating between Asian strains and those from current epidemic. The fact that Peak 4 is barely present in a consensus spectrum of polyproteins from a mixed *flaviviridae* pool including ZIKV and dengue viruses suggests that $f_{RRM}$= 0.334 (601.8nm) is specifically associated to ZIKV activity and must be a subject of future studies of ZF both *in silico* and *in vitro*. The fact that ZIKV envelope and human AXL receptor apparently recognize each other at a sub – harmonic frequency of Peak 4 adds further plausibility to the idea of this peak as a hallmark of ZIKV specific activity.

Based on RRM it is possible to design specific peptides or suggest a low power monochromatic radiation protocol. The availability of a murine model for ZVD provides the opportunity for experimental validation of proposed interventions [20].

## 5 Conclusion

The application of RRM analysis to ZIKV allowed identifying four candidate resonant frequencies that yielded a plausible discrimination between African and Asian/American ZIKV lineages. Association between ZIKV envelope protein and human AXL receptor was supported by RRM, since interaction met all theoretical requirements defined for an interaction between two proteins (sharing a common frequency and resenting opposite phases). Peak 4 at electromagnetic wavelength of λ=601.8nm seems to be the most parsimoniously associated to ZIKV activity. Overall, this study illustrates how a combination of ideas inspired in biophysical modeling and translational bioinformatics can lead to new insights to this serious public health threats.

## 6 Acknowledgements

JLHC expresses his gratitude to Professor Irena Cosic, the designer of the Resonant Recognition Model, for her support and for introducing himinto the application of RRM to the study of infectious diseases. Thanks to Mariana Pita for useful discussion and suggestions.

## 7 Author Contribution

Both authors were included in all phases of preparing this article and final proof reading before sending to print it.

## 8 Conflict of Interest

none declared.

## References

[1] Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. Trans R Soc Trop Med Hyg. 1952; 46: 509–20.

[2] Simpson DIH. Zika virus infection in man. Trans R Soc Trop Med Hyg. 1964; 58: 335–337.

[3] Hayes EB. Zika virus outside Africa. Emerg Infect Dis. 2009; 15: 1347-1350.

[4] Pan American Health Organization / World Health Organization. Zika Epidemiological Update, 16 June 2016. Washington, D.C.: PAHO/WHO; 2016.

[5] Heymann DL, Hodgson A, Sall AA, Freedman DO, Staples JE, Althabe F, et al. Zika virus and microcephaly: why is this situation a PHEIC? The Lancet. 2016; 387: 719-721.

[6] Jin J. Zika Virus Disease. JAMA. 2016; 315: 2482-2489.

[7] Aggarwal R, Aggarwal H, Basu M, Chugh P. Zika virus disease. Int J Community Med Public Health. 2016; 3: 1352-1354.

[8] Fauci SA, Morens DM. Zika Virus in the Americas — yet another arbovirus threat. N Engl J Med. 2016; 374: 601-660.

[9] de Oliveira CS, da Costa Vasconcelos PF. Microcephaly and Zika virus. J Pediatr (Rio J). 2016; 92: 103–105.

[10] http://www.cdc.gov/emailupdates/index.html

[11] Cragan JD, Mai CT, Petersen EE, Liberman RF, Forestieri NE, Stevens AC, et al. Baseline Prevalence of Birth Defects Associated with Congenital Zika Virus Infection —Massachusetts, North Carolina, and Atlanta, Georgia, 2013-2014. Morb Mortal Wkly Rep. 2017; 66: 219–222.

[12] Honein MA, Dawson AL, Petersen EE, Jones AM, Lee EH, Yazdy MM, et al. Birth defects among fetuses and infants of US women with evidence of possible Zika virus infection during pregnancy. JAMA. 2017; 317: 59-68.

[13] https://www.nytimes.com/2016/11/19/health/who-ends-zika-global-health-emergency.html

[14] Murugan NJ, Karbowski LM, Persinger MA. Cosic's Resonance Recognition Model for Protein Sequences and Photon Emission

Differentiates Lethal and Non-Lethal Ebola Strains: Implications for Treatment. Open J Biophys. 2015; 5: 35-43.

[15] Hernandez Caceres JL, Cosic D, Cosic I. Application of the Resonant Recognition Model to the Study of Plasmodium Proteins Involved In Malaria Infection. MD Medical Data. 2015; 7: 007-014.

[16] Cosic I, Hernández Caceres JL, Cosic D. Possibility to interfere with malaria parasite activity using specific electromagnetic frequencies. EPJ Nonlinear Biomed Phys. 2015; 3: 1-11.

[17] Cosic I. Macromolecular Bioactivity: Is It Resonant Interaction between Macromolecules? -theory and applications. IEEE Trans Biomed Eng. 1994; 41: 1101-1114.

[18] Hernández Cáceres JL, Cosic I, Cosic D. Retroviral proteases viewed through the resonant recognition model. MD-Medical Data. 2014; 6: 117-123.

[19] Dotta BT, Murugan NJ, Karbowski M, Lafrenie RM, Persinger MA. Shifting the Wavelengths of Ultraweak Photon Emissions from Dying Melanoma Cells: Their Chemical Enhancement and Blocking Are Predicted by Cosic's Theory of Resonant Recognition Model for Macromolecules. Naturwissenschaften. 2014; 101: 87-94.

[20] Shannan L, Rossi RB, Tesh SR, Azar AE, Muruato KA , Hanley AJ, et al. Characterization of a Novel Murine Model to Study Zika Virus. Am J Trop Med Hyg. 2016; 94: 1362–1369.

[21] Almansour NM, Pirogova E, Coloe PJ, Cosic I, Istivan TS. A bioactive peptide analogue for myxoma virus protein with a targeted cytotoxicityfor human skin cancer in vitro. J Biomed Sci. 2012; 19: 65-74.

[22] Persinger MA. 10–20 Joules as a Neuromolecular Quantum in Medicinal Chemistry: An Alternative Approach to Myriad Molecular Pathways? Curr Med Chem. 2010; 17: 3094-3098.

[23] Logan IS. ZIKA-How fast does this virus mutate? Zool Res. 2016; 37: 110-115.

[24] Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR. Expression Analysis Highlights AXL as a Candidate Zika Virus Entry Receptor in Neural Stem Cells. Cell Stem Cell. 2016; 18: 591-596.

[25] Hernández Cáceres JL. Application of resonant recognition model analysis to Zika virus envelope protein. Rev Electron Biomed / Electron J Biomed. 2015; 3: 15-20.

[26] Veljkovic V, Paessler S. Possible repurposing of seasonal influenza vaccine for prevention of Zika virus infection. F1000Res. 2016; 5: 190.

[27] Cosic I, Cosic D, Lazar K. Tesla, Bioresonances and Resonant Recognition Model. Second International Congress Nikola Tesla - Disruptive innovation. 2017.

[28] Fridell YW, Jin Y, Quilliam LA, Burchert A, McCloskey P, Spizz G, et al. Differential activation of the Ras/extracellular-signal regulated protein kinase pathway is responsible for the biological consequences induced by the Axl receptor tyrosine kinase. Mol Cell Biol. 1996; 16: 135–145.

[29] Tai KY, Shieh YS, Lee CS, Shiah SG, Wu CW. Axl promotes cell invasion by inducing MMP-9 activity through activation of NF-jB and Brg-1. Oncogene. 2008; 27: 4044–4055.

# Improvements in Diabetic Patients' Outcomes in a Clinical Decision Support System

**Sakiko Ota[1*], Kaoru Mogushi[2,3], Aizan Hirai[4], Yoshihito Niimura[1,5] and Hiroshi Tanaka[2,6,7]**

[1] Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

[2] Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

[3] Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, Tokyo, Japan

[4] Chiba Prefectural Togane Hospital, Chiba, Japan

[5] Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

[6] Division of Medical Data Science, Tokyo Medical and Dental University, Tokyo, Japan

[7] Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan

## Abstract

**Background:** There is considerable interest worldwide in prevention and control of diabetes using health information technology systems. Although previous studies have acknowledged improvements of blood pressure control in a clinical decision support system (CDSS), little is known about whether it could contribute to diabetic patients' outcomes.

**Objectives:** To compare diabetic patients' outcomes six months before and after the introduction of a CDSS.

**Methods:** We considered patients with diabetes in a general hospital and ten clinics that embedded a CDSS within an electronic health record in April 2011, and tracked their medical records from October 2010 to March 2012 (split at April 2011; pre- and post-period). The CDSS mainly had two main functions: (1) to visually highlight abnormal laboratory values on longitudinal inspection data; and (2) to provide a list of patients with severe diabetes. Primary outcome measures included glycated hemoglobin (HbA1c), fasting blood glucose, low-density lipoprotein cholesterol (LDL-C), and glomerular filtration rate (eGFR). The values were averaged for each patient and compared between pre- and post-period using the Wilcoxon signed-rank test.

**Results:** A total number of 3,678 blood test results from 705 patients were analyzed. The mean HbA1c levels (and standard deviation) among severe cases (baseline HbA1c ≥ 8.4%) were significantly improved over the study period, from 9.49 ± 1.13% to 8.73 ± 1.23% ($p < 0.001$). In contrast, the moderate group (baseline HbA1c < 8.4%) deteriorated, from 7.61 ± 0.40% to 7.70 ± 0.65%. In terms of fasting blood glucose levels, the severe group (baseline fasting blood glucose level ≥ 160 mg/dl) improved significantly ($p < 0.001$). On the contrary, the moderate group (baseline fasting blood glucose level < 160 mg/dl) deteriorated significantly ($p < 0.001$). The patients' outcomes were improved with the CDSS.

**Conclusion:** Improvements of patients' outcomes were observed, indicating the potential effectiveness of the CDSS. Visually emphasizing abnormal laboratory values and displaying severe patients may be effective for disease control of diabetes patients.

## Keywords

Clinical decision support system; Community network; Diabetes mellitus; Electronic health record; Health information system

**Correspondence to:**

**Dr. Sakiko Ota,**
Graduate School of Health Care Sciences, Tokyo Medical and Dental University1-5-45 Yushima, Bunkyo, Tokyo 113-8510, Japan.
E-mail: osakiko-tky@umin.ac.jp

**EJBI 2018; 14(1):29-36**
Received: December 02, 2017
Accepted: January 09, 2018
Published: January 12, 2018

## 1 Introduction

### 1.1 Background

The number of people with diabetes mellitus has rapidly increased worldwide. The World Health Organization reported that an estimated 422 million adults worldwide were living with diabetes in 2014 [1]. In Japan, the number of patients with diabetes has increased from 2.70 million in 2011 to 3.17 million in 2014, according to a patient survey by the Ministry of Health, Labour and Welfare [2].

Establishing a HIT, such as an electronic health record (EHR) and a clinical decision support system (CDSS), is an important clinical and public health tactic [3, 4, 5, 6, 7, 8, 9, 10]. One potential use of the HIT that has received growing attention is the prevention and control of non-communicable diseases (NCDs) such as diabetes [11, 12, 13, 14, 15].

The Healthcare Information and Management Systems Society (HIMSS) defines clinical decision support as "a process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery" [16]. To date, the CDSSs linked to EHRs play an important part in improving patient safety, efficiency of the healthcare delivery process, and quality of care.

A previous study indicated that the use of a CDSS was associated with improvement of blood pressure control [17]. Moreover, a CDSS integrated with EHRs could moderately improve the outcomes of morbidity [18]. However, little is known about whether CDSSs could contribute to diabetic patients' outcomes [19, 20, 21].

The study aimed to compare patients' diabetes measures six months before and after the introduction of a CDSS. Firstly, we developed a CDSS, which was embedded in an EHR. Secondly, we conducted an experimental study in a general hospital and ten clinics that embedded the CDSS within EHRs.

Figure 1 depicts the structure of the regional EHR, "Wakashio medical network system," which uses the Internet protocol virtual private network (IP-VPN) and the hospital's local area network (LAN) to support collaboration across general practitioners and diabetes specialists in Togane area, Japan. Practitioners and diabetologists logged into the CDSS to see laboratory data (e.g., glycated hemoglobin [HbA1c], fasting blood glucose level, low-density lipoprotein cholesterol [LDL-C], estimated glomerular filtration rate [eGFR]), and comments by other providers when they see patients with diabetes. Figure 2 provides an overview of steps in the referral and counter referral with the CDSS. In the Togane area, diabetologists typically take charge of the patients with severe diabetes and general practitioners manage the patients with moderate diabetes. A local treatment guideline states that patients with diabetes should be referred to diabetologists when their glycated hemoglobin (HbA1c) values exceeded 8.4%. We developed the CDSS to support physicians in noticing the deterioration of HbA1c values and in following the guideline, we developed the CDSS. The CDSS had two functions: (1) visually highlight abnormal laboratory values on longitudinal data (e.g., HbA1c, fasting blood glucose level, low-density lipoprotein cholesterol [LDL-C], and estimated glomerular filtration rate [eGFR]); and (2) to provide a list of patients with severe diabetes (i.e., HbA1c level ≥ 8.4%).

## 1.2 Overview of the "Wakashio Medical Network"

The regional EHR system, known as the "Wakashio Medical Network", was developed at the Chiba Prefectural Togane

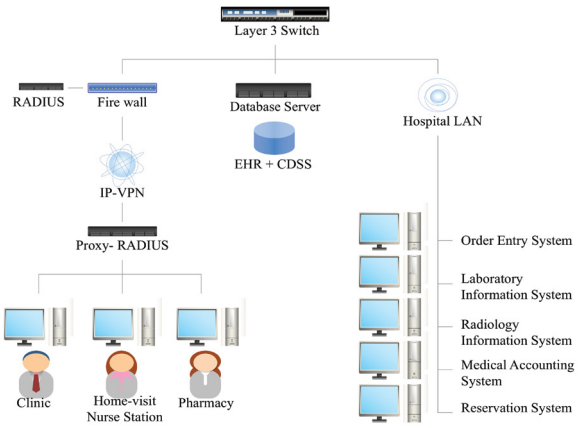



Figure 1: Structure of the Wakashio medical network.
The structure of the medical network system for collaboration across general practitioners and diabetes specialists in the Togane area.

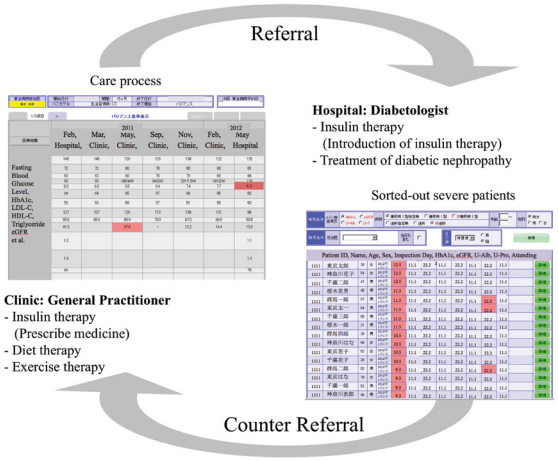



Figure 2: Outline of the medical collaboration with the CDSS.
Overview of the steps in the referral and counter-referral among general practitioners and diabetologists with the CDSS. The main feature of the CDSS is to support general practitioners and diabetologists in noticing abnormal laboratory values.

Hospital, which is a 191-bed local hospital. The shared clinical data was used among general practitioners at 10 clinics and diabetologists at the hospital. The information items were based on Togane's dataset of diabetes care. The dataset included patients' HbA1c, fasting blood glucose, LDL-C, and eGFR.

## 1.3 Features of the CDSS

The CDSS was developed in July 2010 in the EHR stepwise by April 2011. The system was designed in accordance with the clinical guidelines for treating diabetes in Japan [22, 23, 24]. The CDSS mainly had two functions: (1) visually highlight abnormal laboratory values on longitudinal inspection data (e.g., HbA1c, Fasting blood glucose level, and eGFR); and (2) to provide a list of patients with severe diabetes (e.g., HbA1c level ≥ 8.4% [8.0%: Japan Diabetes Society (JDS)]).

Thus, those with severe diabetes had more opportunities to be highlighted in the CDSS.

Both practitioners and diabetologists use the CDSS, such as the table of care process and the sorted-list of severe patients, when performing the examinations. In the table of the care process, it displays the longitudinal inspection data per patient and facilitates the recognition of abnormal laboratory values with pink highlighted color. In the sorted list of severe patients, it displays the basic information of severe patients and abnormal laboratory values are highlighted in pink.

## 2 Materials and Methods

### 2.1 Study Design and Sample Size

We conducted an empirical study in the general hospital and 10 clinics which used the CDSS embedded in EHRs. We compared diabetic patients' lab values six months before and after the introduction of the CDSS. The sample size calculation was made using G*Power version 3 software, which was developed at the Department of Psychology, Heinrich Heine University, Dusseldorf, Germany [25]. Type 1 error [$\alpha = 0.05$ (two-tailed)], power of 0.8, and effect size of 0.3 were used to calculate the minimum sample size. As a result, at least 94 subjects were required. The distribution-based Minimal Clinically Important Difference (MCID) was estimated using the Cohen effect size benchmark. The effect size of 0.3 (i.e., 0.3 SD of the baseline score) indicated an important change and was used as the MCID in this study.

### 2.2 Data Source

We used data from EHRs in the Chiba Prefectural Togane Hospital, Japan. Participants in the study were selected based on the following criteria: (1) persons aged 20 years old or over; and (2) whose baseline of HbA1c level ≥ 6.9% (6.5% JDS). Baseline or pre-intervention data were collected from October 1, 2010 to March 31, 2011. Twelve-month follow-up or post-intervention data were collected from October 1, 2011 to March 31, 2012. We excluded data between April 1 and September 30, 2011 to assess the impact on HbA1c levels, which identifies the average plasma glucose level over the preceding few months [26]. We conducted the study based on the 962 consecutively treated patients with diabetes between October 1, 2010 and March 31, 2012 in Togane area, Japan. Patients were excluded if they had laboratory tests fewer than two times during the pre- or the post- period. As a result, 257 patients were excluded, and data of the remaining 705 patients and a total of 3,678 blood test records were then analyzed.

We used the HbA1c measurement based on the National Glycohemoglobin Standardization Program (NGSP). The NGSP is routinely applied in many other countries [22, 27, 28, 29]. In Japan, the HbA1c measurement based on the Japan Diabetes Society (JDS) had been used until March 31, 2012 and the NGSP has been used since April 1, 2012. The conversion equation from the JDS to NGSP is as follows: NGSP (%) = JDS (%) + 0.4% [22, 30].

### 2.3 Analysis

Primary outcome measures included HbA1c, fasting blood glucose, LDL-C, and eGFR. These were averaged for each patient within the pre- and post- periods, and compared using the Wilcoxon signed-rank test. First, we analyzed the overall data. Second, we compared the data among severity-stratified groups. We stratified the patients into two groups, based on the baseline, into severe (e.g., HbA1c level ≥ 8.4% [8.0% JDS]) and moderate groups (e.g., HbA1c level < 8.4% [8.0% JDS]). Furthermore, we analyzed correlations between pre-intervention (baseline) and difference during pre- and post- intervention with the Pearson's correlation coefficient. We adapted a 5%, two-tailed significance level. R 3.0.2 software (R Foundation for Statistical Computing, Vienna, Austria) was used for all statistical analyses. As for the participant consent, the documentation of informed consent was waived by the institutional review board (IRB) of Tokyo Medical and Dental University. Based on ethical guidelines for epidemiological research presented by the Ministry of Health, Labour and Welfare, we employed an opt-out methodology to obtain the participants. The study protocol and informed consent procedure were approved by the institutional review boards of Tokyo Medical and Dental University.

Table 1: Patient characteristics of HbA1c levels before the introduction.

| Characteristics | Mean HbA1c levels in the pre- period | | | |
|---|---|---|---|---|
| | < 8.4 %¶ (n=470) | | ≥ 8.4 %¶ (n=235) | |
| | Median (IQR) | | Median (IQR) | |
| Age | 67 | (60–76) | 64 | (55–73) |
| Sex* (Female) | 276 | (58.7) | 138 | (58.7) |
| Fasting blood glucose level (mg/dl) | 169 | (148–194) | 217 | (182–260) |
| eGFR (mL/min/1.73m$^2$) | 73.2 | (61.2–88.3) | 81.8 | (62.4–96.6) |
| LDL-C (mg/dl) | 95 | (77–116) | 98 | (84–119) |

HbA1c, glycated hemoglobin; eGFR, estimated glomerular filtration rate; LDL-C, low-density lipoprotein cholesterol; IQR, interquartile range. ¶HbA1c measurement based on the National Glycohemoglobin Standardization Program (NGSP). *Sex is showed as number and percentage.

# 3 Results

## 3.1 Baseline Characteristics

Baseline characteristics are shown in Table 1. Briefly, 55.5% of patients were ≥ 65 years old and 26.8% of patients were ≥ 75 years old. There were 414 female patients (58.7%). The laboratory parameters are also listed in Table 1. Of 705 patients, 235 patients (33.3%) had a baseline HbA1c level of ≥ 8.4% (8.0% JDS). In addition, 497 patients (70.5%) had baseline fasting blood glucose levels ≥ 160 mg/dl. Forty-nine patients (7.0%) had baseline LDL-C levels ≥ 140 mg/dl, and 156 patients (22.1%) had baseline eGFR<60 mL/min/1.73m$^2$.

## 3.2 Patient Outcomes

A total number of 3,678 blood test results from 705 patients were analyzed. The mean of the test frequency per

Table 2: Comparison diabetic patients' outcomes six months before and after the introduction of the CDSS.

| Laboratory values | Pre-introduction | Post-introduction | *P* value* | Cohen's *d* |
|---|---|---|---|---|
| **HbA1c (%)** | | | | |
| All (n = 705) | | | < 0.001‡ | 0.175 |
| Mean ± SD | 8.23 ± 1.15 | 8.04 ± 1.01 | | |
| Median (IQR) | 7.90 (7.45–8.68) | 7.77 (7.33–8.50) | | |
| < 8.4 NGSP (8.0 JDS) (n = 470) | | | 0.025† | 0.158 |
| Mean ± SD | 7.61 ± 0.40 | 7.70 ± 0.65 | | |
| Median (IQR) | 7.60 (7.27–7.90) | 7.55 (7.20–7.98) | | |
| ≥ 8.4 NGSP (8.0 JDS) (n = 235) | | | < 0.001‡ | 0.642 |
| Mean ± SD | 9.49 ± 1.13 | 8.73 ± 1.23 | | |
| Median (IQR) | 9.09 (8.68–9.95) | 8.61 (7.87–9.20) | | |
| **Fasting blood glucose level(mg/dl)** | | | | |
| All (n = 705) | | | 0.28 | 0.032 |
| Mean ± SD | 191.7 ± 49.8 | 190.1 ± 51.0 | | |
| Median (IQR) | 180.0 (155.0–214.7) | 178.6 (151.5–216.5) | | |
| < 160 (n = 208) | | | < 0.001‡ | 0.613 |
| Mean ± SD | 144.2 ± 9.2 | 165.2 ± 37.9 | | |
| Median (IQR) | 144.7 (137.0–152.0) | 153.6 (141.5–178.6) | | |
| ≥ 160 (n = 497) | | | < 0.001‡ | 0.224 |
| Mean ± SD | 211.6 ± 46.2 | 200.5 ± 52.2 | | |
| Median (IQR) | 199.3 (177.5–231.0) | 189.8 (161.5–231.0) | | |
| **eGFR (ml/min/1.73 m$^2$)** | | | | |
| All (n = 687) | | | < 0.001‡ | 0.097 |
| Mean ± SD | 76.8 ± 25.6 | 74.3 ± 25.8 | | |
| Median (IQR) | 75.6 (61.6–92.1) | 73.4 (58.9–89.2) | | |
| ≥ 60 (n = 531) | | | < 0.001‡ | 0.874 |
| Mean ± SD | 85.9 ± 20.7 | 83.0 ± 21.5 | | |
| Median (IQR) | 82.7 (71.3–95.7) | 80.0 (68.8–92.8) | | |
| < 60 (n = 156) | | | 0.004† | 0.064 |
| Mean ± SD | 45.6 ± 13.1 | 44.7 ± 15.0 | | |
| Median (IQR) | 50.3 (39.6–55.4) | 47.2 (37.5–55.6) | | |
| **LDL-C (mg/dl)** | | | | |
| All (n = 635) | | | < 0.001‡ | 0.148 |
| Mean ± SD | 99.7 ± 29.2 | 95.4 ± 28.8 | | |
| Median (IQR) | 97.0 (78.7–117.0) | 93.0 (75.5–114.0) | | |
| < 140 (n = 586) | | | < 0.001‡ | 0.120 |
| Mean ± SD | 94.5 ± 22.9 | 91.9 ± 20.3 | | |
| Median (IQR) | 94.3 (77.0–112.5) | 91.0 (74.0–110.0) | | |
| ≥ 140 (n = 49) | | | < 0.001‡ | |
| Mean ± SD | 161.8 ± 23.3 | 136.2 ± 26.7 | | |
| Median (IQR) | 155.4 (145.3–167.5) | 136.8 (119.6–152.0) | | |

HbA1c, glycated hemoglobin; eGFR, estimated glomerular filtration rate; LDL-C, low-density lipoprotein cholesterol; NGSP, National Glycohemoglobin Standardization Program; JDS, Japan Diabetes Society; SD, standard deviation; IQR, interquartile range. *Wilcoxon signed rank test; † p-value < 0.05; ‡ p-value < 0.001.

person was 5 (3–6) times during the pre- and the post- periods. Results are given for the group as an overall and stratified by severe and moderate cases. Continuous variables are summarized as mean ± standard deviation (SD) and median with interquartile range (IQR) values. These results are shown in Table 2.

The mean HbA1c levels among severe cases (baseline ≥ 8.4%) improved significantly over the study period, from 9.49 ± 1.13% to 8.73 ± 1.23% (p<0.001, Cohen's $d$ = 0.642). In contrast, the moderate group (baseline <8.4%) deteriorated significantly (p = 0.025, Cohen's $d$ = 0.158), from 7.61 ± 0.40% to 7.70 ± 0.65%. Because the effect size of 0.3 (i.e., 0.3 SD of the baseline score) was used as the MCID in this study, the difference was not important clinically. In terms of fasting blood glucose levels, the severe group (baseline ≥ 160 mg/dl) improved significantly (p < 0.001, Cohen's $d$ = 0.224), from 211.6 ± 46.2 mg/dl to 200.5 ± 52.2 mg/dl. On the contrary, the moderate group (baseline < 160 mg/dl) significantly deteriorated, from 144.2 ± 9.2 mg/dl to 165.2 ± 37.9 mg/dl (p < 0.001, Cohen's $d$ = 0.613). Figure 3 depicts correlations between the mean HbA1c in the pre- period and the difference of HbA1c during the pre- and the post- period. The Pearson's correlation coefficient was –0.56 (95% confidence interval [CI] –0.60 to –0.50, p < 0.001). The greater HbA1c level in the pre-period was associated with a greater reduction. S1 and S2 Figures show equivalence tests for comparing correlation coefficients between all patients and patients outside the boundary (i.e., those whose HbA1c levels were 8.4 ± 0.3% in the pre- period). In the severe group, the correlation coefficients of all patients and patients outside the boundary were –0.47 (95%CI –0.56 to –0.36) and –0.42 (95%CI –0.53 to –0.29), respectively (ρ = –0.45). As for the moderate group, the correlation coefficients of all patients and patients outside the boundary were –0.24 (95%CI –0.32 to –0.15) and –0.18 (95%CI –0.28 to –0.08), respectively (ρ = –0.21).

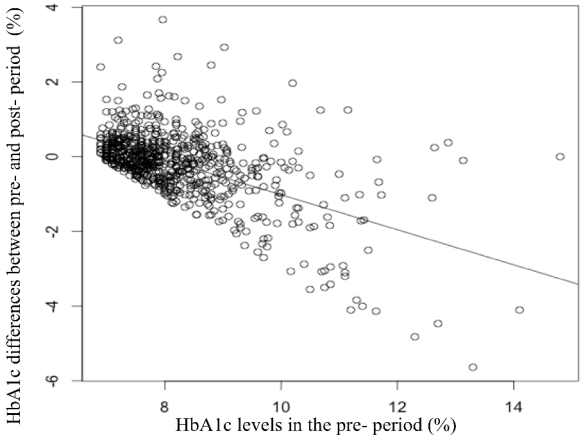
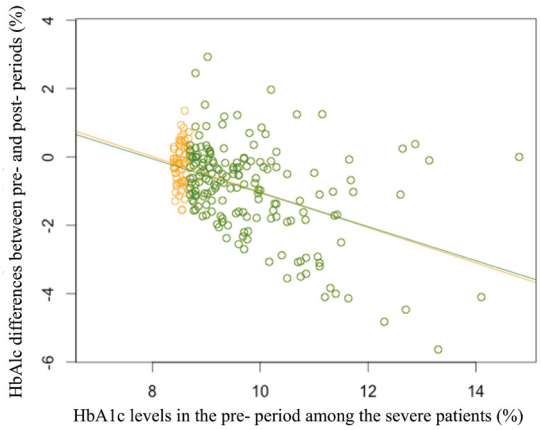



Figure 3: Correlations between the baseline HbA1c and the difference of HbA1c during pre- and post- period.

Correlations between the HbA1c in the pre- period and the difference in HbA1c between the pre- and the post- period. The Pearson's correlation coefficient is –0.56 (95% CI –0.60 to –0.50, p < 0.001).




S1 Figure: Correlations between the HbA1c in the pre- period and the difference of HbA1c during the pre- and the post-periods among the severe patients.

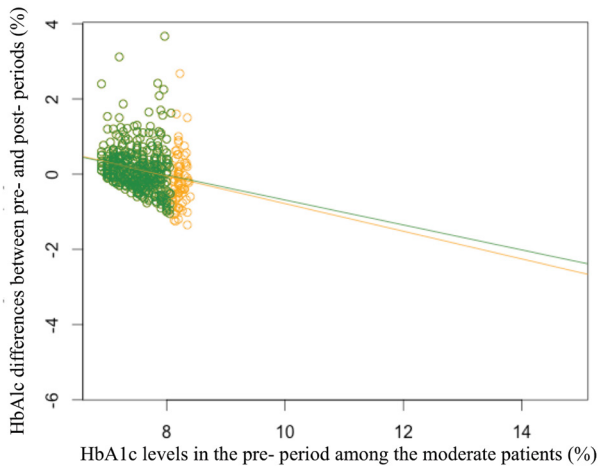
Correlations between the HbA1c in the pre- period and the difference in HbA1c between the pre- and the post- period among the severe patients. The correlation coefficients of all patients and patients outside the boundary (i.e., 8.4% ≤ HbA1c <8.7% in the pre- period) were –0.47 (95%CI –0.56 to –0.36) and –0.42 (95%CI –0.53 to –0.29), respectively (ρ = –0.45).

## 4 Discussion

### 4.1 Diabetic Patients' Outcomes with the CDSS

To compare diabetic patients' outcomes six months before and after the introduction of the CDSS, we considered 705 patients in the general hospital and ten clinics that embedded the CDSS within EHRs in April 2011, and tracked their medical records from October 2010 to March 2012 (split at April 2011; pre- and post- period). A total number of 3,678 blood test results were analyzed. The mean HbA1c levels among severe cases (HbA1c ≥ 8.4% in the pre- period) were significantly improved over the study period (p < 0.001). In contrast, the moderate group (HbA1c < 8.4% in the pre- period) deteriorated significantly (p = 0.025), but the difference was not important clinically.

Improvements of patients' outcomes were observed, indicating the potential effectiveness of the CDSS. The CDSS mainly had two functions: (1) visually highlight abnormal laboratory values on longitudinal inspection data; and (2) to provide a list of patients with severe diabetes. In the empirical study, HbA1c values were highlighted when the values exceeded 8.4% based on the treatment guideline. Therefore, the values of the severe patients (HbA1c ≥ 8.4% in the pre- period) were more often highlighted than the moderate patients (HbA1c < 8.4% in the pre- period). Through repeated notifications, the CDSS has more opportunity to improve patients' outcomes of diabetes care. According to a study in US, diabetes care practice often does not achieve recommended diabetes care standards [31]. In that case, the repeated notifications may contribute to treatment decisions and to increase adherence to recommended diabetes care, although the mechanism remains unclear. Through interviews, some practitioners stated that the overlooking of deterioration of




S2 Figure: Correlations between the HbA1c in the pre- period and the difference of HbA1c during the pre- and the post- periods among the moderate patients.

Correlations between the HbA1c in the pre- period and the difference in HbA1c between the pre- and the post- period among the moderate patients. The correlation coefficients of all patients and patients outside the boundary (i.e., 8.1% ≤ HbA1c < 8.4% in the pre- period) were −0.24 (95%CI −0.32 to −0.15) and −0.18 (95%CI −0.28 to −0.08), respectively (ρ = −0.21).

inspection items was almost eliminated and it became easier for practitioners to follow the treatment guideline. Further research into how the repeated notifications contribute to treatment decisions is needed to establish stronger evidence.

We also compared the two groups (i.e., severe and moderate groups) on the correlation between the mean HbA1c in the pre-period and the difference of HbA1c during the pre- and the post-period. Pearson's product-moment correlations in the severe and moderate groups were −0.47 and −0.24 respectively. Furthermore, we conducted an equivalence test for comparing correlation coefficients between the severe and moderate groups ($p < 0.001$). In the severe group, whose values were highlighted more often by the CDSS, the HbA1c values significantly improved, so the CDSS might contribute to improve patients' outcomes. However, there was the possibility of regression to the mean because of a lack of control group. In a future round of the empirical study, we plan to quantify the effect of the CDSS with a control group.

Previous studies shown mixed effects on patients' outcomes after the introduction of a CDSS. In a randomized trial in primary care clinics in Canada, individualized electronic decision support and reminders improved diabetes care [12]. The mean follow-up was 5.9 months and the main outcomes were HbA1c, blood pressure, and LDL-C. The CDSS improved the process of care and some clinical markers of diabetes care. In US, several studies of diabetes care with a CDSS have been conducted during the past two decades. One of the studies was conducted in primary care and examined three outcomes (i.e., HbA1c, LDL-C, and blood pressure) at the 6-month follow-up [32]. The research found that the CDSS significantly improved HbA1c and some aspects of blood pressure. On the other hand, a review paper suggested that the use of a CDSS resulted in only modest improvements in quality of diabetes care and further empirical studies were needed [5]. To date, very little experimental work has been done on the effectiveness of a CDSS and little is known about whether a CDSS could contribute to patients' outcomes. This study contributes important insights about the effectiveness of a CDSS on diabetes care in Japan.

### 4.2 Limitations of the Study

Several limitations of this study should be acknowledged. First, the results with the empirical study are generalizable only to similar settings. Although similar evaluation methods are used in many settings in software engineering and medical informatics literature [33, 34], setting up a control group would have improved the quality of this study. Second, there is the possibility that observed changes represented regression toward the mean because of the lack of a control group. Against the interpretation, note that we used several laboratory values to compute the mean of HbA1c levels in the periods. The mean of the test frequency per person was 5 (3–6) times during the pre- and the post- periods. Furthermore, to reduce measurement error, we conducted equivalence tests for comparing correlation coefficients (i.e., HbA1c levels were 8.4 ± 0.3% in the pre- period). From these results, the population correlation coefficients of all patients and patients outside the boundary were not different. Third, there is a possibility of other factors that may affect patients' outcomes (e.g., changes of treatment plan with the passage of time) and the Hawthorne effect. Despite these limitations, the findings of this study provide useful information regarding the advantages of using the CDSS in clinical practice.

## 5 Conclusion

To compare diabetic patients' outcomes six months before and after the introduction of the CDSS, we studied 705 patients in the general hospital and ten clinics that embedded the CDSS within EHRs in April 2011, and tracked their medical records from October 2010 to March 2012. After the introduction of the CDSS, the mean HbA1c levels among severe cases were significantly improved over the study period ($p < 0.001$). In contrast, the moderate group deteriorated significantly ($p = 0.025$). Improvements of patients' outcomes were observed, indicating the potential effectiveness of the CDSS. Visually emphasizing abnormal laboratory values and displaying severe patients may be an effective way for disease control of diabetes patients.

## 6 Acknowledgement

Gratitude is expressed to the staff at the Chiba Prefectural Togane Hospital for providing and collecting the laboratory data, and to Tomoyasu Tanaka at NTT DATA Corporation for technical assistance.

## 7 Author Contribution

All others have contributed substantially to the conception and design, occasion of data, or analysis; S. Ota has drafted the article; S. Ota and A. Hirai performed the data collection of participants; S. Ota and K. Mogushi performed the statistical analysis for the patient outcomes; S. Ota, K. Mogushi, and A. Hirai performed the interpretation the data; K. Mogushi, Y. Niimura, and H. Tanaka revised the article critically for important intellectual content; All authors have given approval of the final version to be published.

## 8 Financial Disclosure

This work was supported by Health Labour Sciences Research Grant project for research on region medical: matching fund subsidy from Ministry of Health Labour and Welfare in Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 9 Competing Interests

The authors have declared that no competing interests exist.

## References

[1] World Health Organization. Global report on diabetes. Geneva: WHO Press; 2016.

[2] http://www.mhlw.go.jp/english/database/db-hss/sps_2014.html

[3] Coiera E. Guide to health informatics. 3rd ed. United States: CRC Press; 2015.

[4] Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. Int J Med Inform. 2014; 83: 779-796.

[5] Ali SM, Giordano R, Lakhani S, Walker DM. A review of randomized controlled trials of medical record powered clinical decision support system to improve quality of diabetes care. Int J Med Inform. 2016; 87: 91-100.

[6] Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. BMC Med Inform Decis Mak. 2012; 12:10.

[7] Wu RC, Straus SE. Evidence for handheld electronic medical records in improving care: a systematic review. BMC Med Inform Decis Mak. 2006; 6: 26.

[8] Smith AJ, Skow A, Bodurtha J, Kinra S. Health information technology in screening and treatment of child obesity: a systematic review. Pediatrics. 2013; 131: e894-902.

[9] World Health Organization. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. Geneva: WHO Press; 2011.

[10] World Health Organization. Legal frameworks for eHealth: based on the findings of the second global survey on eHealth. Geneva: WHO Press; 2012.

[11] Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic health records and quality of diabetes care. N Engl J Med. 2011; 365: 825-33.

[12] Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D, et al. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ. 2009; 181: 37-44.

[13] Crosson JC, Ohman-Strickland PA, Cohen DJ, Clark EC, Crabtree BF. Typical electronic health record use in primary care practices and the quality of diabetes care. Ann Fam Med. 2012; 10: 221-227.

[14] Balas EA, Krishna S, Kretschmer RA, Cheek TR, Lobach DF, Boren SA. Computerized knowledge management in diabetes care. Med Care. 2004; 42: 610-621.

[15] International Telecommunication Union. Be He@lthy, Be Mobile. Geneva: 2014.

[16] Osheroff J, Teich J, Levick D, Saldana L, Velasco F, Sittig D, et al. Improving Outcomes with Clinical Decision Support: An Implementer's Guide. 2nd ed. HIMSS; 2012.

[17] Samal L, Linder JA, Lipsitz SR, Hicks LS. Electronic health records, clinical decision support, and blood pressure control. Am J Manag Care. 2011; 17: 626-632.

[18] Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health. 2014; 104: e12-e22.

[19] Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc. 2011; 18: 327-34.

[20] Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA. 2005; 293: 1223-1238.

[21] Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med. 2006; 144: 742-752.

[22] Seino Y, Nanjo K, Tajima N, Kadowaki T, Kashiwagi A, Araki E, et al. Report of the committee on the classification and diagnostic criteria of diabetes mellitus. J Diabetes Investig. 2010; 1: 212-228.

[23] The Japan Diabetes Society. Treatment Guide for Diabetes 2007. Tokyo: Bunkodo; 2007.

[24] The Japan Diabetes Society. Evidence-based Practice Guideline for the Treatment for Diabetes in Japan 2013. Tokyo: Nunkodo; 2013.

[25] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007; 39: 175-91.

[26] Nathan DM, Turgeon H, Regan S. Relationship between glycated haemoglobin levels and mean glucose levels over time. Diabetologia. 2007; 50: 2239-2244.

[27] World Health Organization. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation. Geneva: WHO Press; 2011.

[28] Amemiya S, Hoshino T. The worldwide standardization of hemoglobin A1c measurement: the current statement from the Japan Diabetes Society and the issues to be solved. Rinsho Byori. 2013; 61: 594-601.

[29] Kashiwagi A. Development of the HbA1c international standardization of HbA1c measurement in Japan Diabetes Society. Rinsho Byori. 2013; 61: 585-593.

[30] Kashiwagi A, Kasuga M, Araki E, Oka Y, Hanafusa T, Ito H, et al. International clinical harmonization of glycated hemoglobin in Japan: From Japan Diabetes Society to National Glycohemoglobin Standardization Program values. J Diabetes Investig. 2012; 3: 39-40.

[31] McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003; 348: 2635-2645.

[32] O'Connor PJ, Sperl-Hillen JM, Rush WA, Johnson PE, Amundson GH, Asche SE, et al. Impact of electronic health record clinical decision support on diabetes care: a randomized trial. Ann Fam Med. 2011; 9: 12-21.

[33] Kampenes VB, Dyba T, Hannay JE, Sjoberg DIK. A systemetic review of quasi-experiments in software engineering. Information and Software Technology. 2009; 51: 71-82.

[34] Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc. 2006; 13: 16-23.

# Medication Pattern Mining Considering Unbiased Frequent Use by Doctors

**Yuji Morita[1], Masatoshi Yoshikawa[1*], Noboru Kada[1], Akihiro Hamasaki[2], Osamu Sugiyama[3], Kazuya Okamoto[4] and Tomohiro Kuroda[4]**

[1] Graduate School of Informatics, Kyoto University, Kyoto, Japan

[2] Center for Diabetes & Endocrinology, Kitano Hospital, Osaka, Japan

[3] Preemptive Medical and Lifestyle Disease Research Center, Kyoto University Hospital, Kyoto, Japan

[4] Division of Medical IT & Administration Planning, Kyoto University Hospital, Kyoto, Japan

## Abstract

**Background:** Many previous studies on mining prescription sequences are based only on frequency information, such as the number of prescriptions and the total number of patients issued the prescription. However, in cases where a very small number of doctors issue a prescription representative of a certain medication pattern to many patients many times, the prescribing intention of this very small number of doctors has a great influence on pattern extraction, which introduces bias into the final extracted frequent prescription sequence pattern.

**Objectives:** We attempt to extract frequent prescription sequences from more diverse perspectives by considering factors other than frequency information to ensure highly reliable medication patterns.

**Methods:** We propose the concept of unbiased frequent use by doctors as a factor in addition to frequency information based on the hypothesis that a prescription used by many doctors unbiasedly is a highly reliable prescription. We propose a medication pattern mining method that considers unbiased frequent use by doctors. We conducted an evaluation experiment using indicators based on clinical laboratory test results as a comparative evaluation of the existing method, which relied only on frequency, and included consideration of unbiased frequent use by doctors by the proposed method.

**Results:** The weighted average value of the top $k$ for two different evaluation methods is obtained.

**Conclusions:** The study suggested that our medication pattern mining method considering unbiased frequent use by doctors is useful in certain situations such as when the clinical laboratory test value is outside of the normal value range.

## Keywords

Electronic medical record; Clinical decision support; Unbiased frequent use by doctors; Sequential pattern mining

**Correspondence to:**

**Masatoshi Yoshikawa**
Department of Social Informatics, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo, Kyoto, Japan.
E-mail: yoshikawa@i.kyoto-u.ac.jp

**EJBI 2018; 14(1):37-44**
Received: December 20, 2017
Accepted: January 22, 2018
Published: January 29, 2018

## 1 Introduction

Electronic medical records (hereafter called "EMRs") are designed for the primary use of data, such as the efficiency of medical work by order entry systems and medical accounting systems. Secondary data, such as medical research, decision support, and educational support for young doctors, are expected to accumulate via promoting the computerization of medical data. Especially, computer-based clinical decision support systems (hereafter called "CDSS") to support doctor's clinical decision making is an important example of secondary use of EMRs [1]. The CDSS is a system that provides better medical care to patients by reducing decision mistakes and sharing medical evidence when medical staff makes decisions, such as diagnosis and prescription [2].

Clinical decision support in primary care has attracted attention recently, and many previous studies have aimed to improve the quality of medical care by presenting appropriate treatment procedures to doctors using sequential data mining [3]. For example, when conducting medication therapy (one of the therapies used in primary care), the doctor appropriately selects the type and dose of the next prescription medicine based on the relationship among patient's specific information, the current medication and the next medication. Then, many clinical experiences can be shared, and the next prescription medicine can be efficiently decided within a range of adaptations by suggesting frequent sequences extracted from prescription data under the same conditions to the doctor as well as potential medications that can be prescribed. Many previous studies have extracted prescription sequences based only on frequency information, such as the number of prescriptions and the total number of patients issued the prescription [4, 5, 6].

However, simply applying sequential data mining techniques based on frequency information to EMRs are not always appropriate. Because each doctor has his/her own strategy for medication, consideration of difference among medication patterns of doctors is necessary. In cases where a very small number of doctors issue a prescription representative of a medication pattern to many patients many times, the prescribing intention of this very small number of doctors has a large influence on pattern extraction. This problem introduces bias to the final extracted prescription sequences. Therefore, we attempt to extract prescription sequences from more diverse perspectives by considering unbiased frequent use by doctors as an additional factor. Our hypothesis is that prescriptions performed by many doctors unbiasedly are highly reliable.

## 2 Methods

### 2.1 Formulation of Frequent Pattern Extraction Considering Unbiased Frequent Use by Doctors

We start with definition of basic terms. A *prescription* is a pair of medicine and dose such as (thiamazole, 5 mg). A *sequence* (*pattern*) is a consecutive sequence of two prescriptions such as (thiamazole, 5 mg) → (thiamazole, 10 mg). We say a doctor performed a sequence when the doctor performed the prescription which appears in the consequent of the sequence when the previous prescription was the antecedent of the sequence. For example, let $\alpha$ be the sequence (thiamazole, 5 mg) → (thiamazole, 10 mg). If a doctor performed the prescription (thiamazole, 10 mg) when the previous prescription was (thiamazole, 5 mg), we say the doctor performed $\alpha$.

We propose the concept that the unbiased frequent use by doctors is a factor in addition to frequency information. Thus, we formulated a prescription sequence mining method that considered unbiased frequent use by doctors.

More specifically, we propose the DF-score as a new score for a sequence calculated by the following formula based on the *Support* and D-rank value.

$$DF-score(\alpha)=Support(\alpha)\times D-rank(\alpha)$$

Here, α is a sequence, $Support(\alpha)$ is α's support value, and $D-rank(\alpha)$ is the value of unbiased frequency of α performed by doctors. In other words, the $DF-score(\alpha)$ is an index that combines information based on the patient-based frequency and the doctors' unbiased usage.

We define the unbiased frequent use by doctors as an indicator based on the following factors.

Rate of doctors who performed a certain medication pattern: $D_{rate}$

The unbiased variance representing the usage frequency dispersion for a certain prescription sequence: $V_{UB}(\alpha)$

Then, the unbiased frequent rank of α is calculated by the following formula:

$$D-rank(\alpha)=\frac{D_{rate}(\alpha)}{V_{UB}(\alpha)}$$

$D_{rate}(\alpha)$ represents the rate of doctors who performed a sequence α and is defined as follows:

$$D_{rate}(\alpha)=\frac{count(D(\alpha))}{|D|}$$

$count(D(\alpha))$ indicates the number of doctors who performed sequence, $\alpha$ and $|D|$ indicates the total number of all doctors. Thus, $D_{rate}$ represents the rate of doctors who have performed sequence α out of all doctors.

$V_{UB}(\alpha)$ is defined as the unbiased variance representing the usage frequency dispersion for sequence $\alpha$ for each doctor as follows:

$$V_{UB}(\alpha)=\frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2$$

Here, let $\sum=\{D_1,D_2,\ldots,D_n\}$ be the set of all doctors. The proportion of patients who are prescribed sequence $\alpha$ out of all prescribed patients by doctor $D_i$ is $d_i$, and $\bar{d}$ shows the average value of the proportion of patients prescribed by all doctors using sequence $\alpha$. We obtain $D-rank(\alpha)$ by dividing $D_{rate}(\alpha)$ by $V_{UB}(\alpha)$. Therefore, a sequence performed by many doctors unbiasedly yields higher $D-rank(\alpha)$ value.

Based on these calculations, Figure 1 shows the entire process from data acquisition to calculation of the DF-score of prescriptions. The procedure is as follows:

1 A correspondence table showing the relationship between the product name and the generic name of medication is prepared.

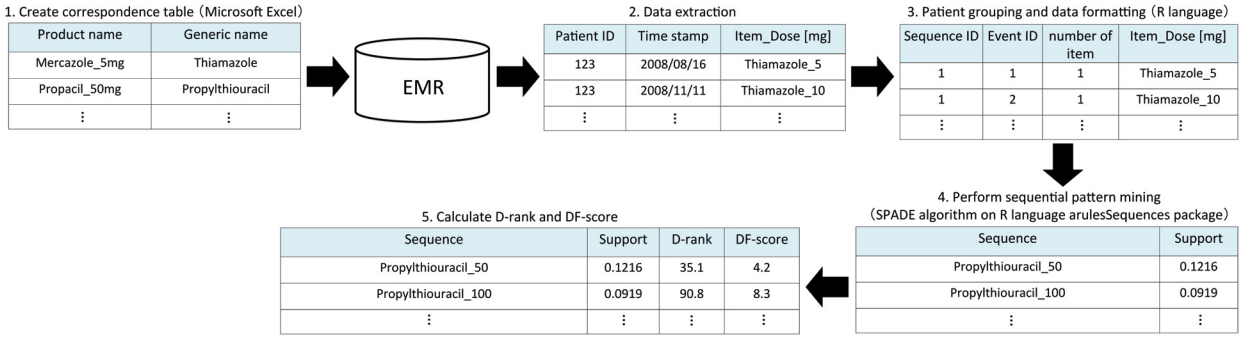



Figure 1: Calculation procedure for unbiased frequent use by doctors.

2 The prescription data and clinical laboratory test value data from the patients are extracted from the EMRs for analysis based on the prepared correspondence table.

3 To preprocess the data for sequential pattern mining, the data extracted from the EMRs is converted into a transaction form after grouping each patient based on the clinical laboratory test result.

4 Sequential pattern mining is applied to the transformed data set to extract frequent patterns.

5 The unbiased frequent use by doctors (D-rank) and the DF-score of each extracted frequent pattern are calculated. Finally, frequent prescription patterns are extracted based on the DF-score.

# 3 Experiment

## 3.1 Data Set

The information in the EMR includes the patient's name, age, progress record, and medical imaging. Under the approval of ethical committee of Kyoto University School of medicine, we applied prescription data and clinical laboratory test value data from patients with thyroid disease. These data sets contain anonymized data acquired from September 20, 2000, to December 22, 2015, by the EMR system KING (Kyoto University Hospital Information Galaxy) used at Kyoto University Hospital.

## 3.2 Evaluation Method

To evaluate the objective effectiveness of the extracted medication pattern, we assessed the degree to which the clinical laboratory test values improved after prescribing the medication pattern. For example, when the clinical laboratory test values were individually considered on a case-by-case basis, some cases improved and others deteriorated. On this occasion, countless evaluation approaches can be used depending on the aspect to be interpreted from the clinical laboratory test value (i.e., defining improvement and deterioration based on different criteria). Therefore, we classify individual cases into several groups and propose an evaluation method from the perspective of the improvement rate of clinical laboratory test values for each group. For this purpose, we propose the following two evaluations.

- Evaluation method 1: An evaluation method based on the normal value range.
- Evaluation method 2: An evaluation method based on the median value of the normal range.

We describe specific definitions and procedures for the above two evaluation methods. First, we explain evaluation method 1. In medication therapy for many diseases, including thyroid disease, prescribed medication is generally selected in consideration of maintaining the clinical laboratory test values within the normal value range as much as possible. Therefore, the normal value range (which is the medical standard) is considered very useful information for the measurement of the improvement of the clinical laboratory test value. Evaluation method 1 is a method based on this normal value range. A specific score is assigned to each case based on the clinical laboratory test value state, such as a transition of one, and the final evaluation value is calculated by averaging the scores of all cases. The method used to assign the score is defined based on the clinical opinion of the specialist. Figure 2 shows an example of a method used to assign a score according to each clinical laboratory test value state. The method gives a score when the examination value state at the time of prescription is within the normal value range (Figure 2a), has an abnormal value (Low) (Figure 2b) or has an abnormal value (High) (Figure 2c). The transition of the clinical laboratory test value after the prescription of a patient in a certain clinical laboratory test value state in which a medication pattern is performed is classified as described. Then, a score is assigned for each case according to the classification result. Let the clinical laboratory test value at prescription be $V$. The next clinical laboratory test value of $V$ will be $V'$, the upper limit value of the normal value range will be $N_U$, and the lower limit value of the normal value range will be $N_L$.

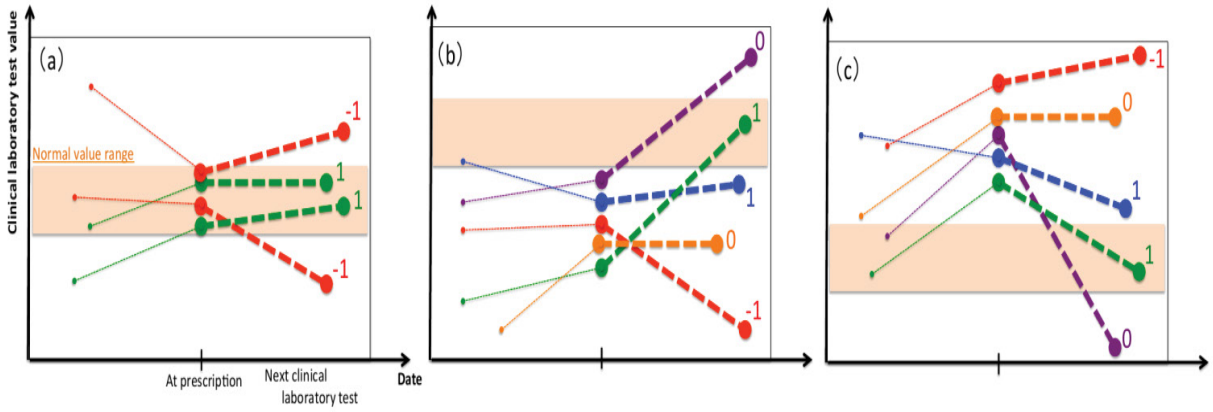



Figure 2: Examples of a score assignment method in the evaluation method 1.

- Improvement
- $N_L \leqq V' \leqq N_U$
- $V < V' < N_L$
- $N_U < V' < V$

⇒ Assign score "1"

- Remain unchanged
- $(N_L > V) \wedge (V = V')$
- $(N_U < V) \wedge (V = V')$

⇒ Assign score "0"

- Exceedance
- $(N_L > V) \wedge (V = V')$
- $(N_U < V) \wedge (N_L > V')$

⇒ Assign score "0"

- Deleterious change
- $(N_L \leqq V \leqq N_U) \wedge (N_U < V')$
- $(N_L \leqq V \leqq N_U) \wedge (N_L > V')$
- $(N_L > V) \wedge (V > V')$
- $(N_U < V) \wedge (V < V')$

⇒ Assign score "-1"

The final evaluation value is calculated by the following formula based on the score assigned to each case.

$$\frac{1}{n}\sum_{i=1}^{n}(Score_i)$$

$Score_i$ indicates the score value in case *i*. The final evaluation value calculated by the above formula represents the average value of the scores of $n$ cases. Evaluation method 1 interprets the pattern with the final evaluation value closer to 1 as the medication pattern that results in better improvement of the clinical laboratory test value.

Second, we explain evaluation method 2. Since the above-described evaluation method 1 is a method based only on the normal value range, concrete values from the clinical laboratory test values are not taken into consideration. Therefore, even though the score values assigned in evaluation method 1 are the same, the actual clinical laboratory test values may differ. Conversely, evaluation method 2 focuses on the median of the normal range using a different approach than evaluation method 1. We compare quantitative values between case groups before and after prescription using the distance from the median of the normal range. We evaluate the transition of the clinical laboratory test value of each case by RMS (Root Mean Square) using the median of the normal range. A large weight can be given to a case located at a distance from the median of the normal range using RMS, and the influence of the deteriorated case is considered to more appropriately reflect the final evaluation value. The evaluation formula is defined as follows:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(d'_i)^2} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i)^2}$$

$d_i$ is the difference between the clinical laboratory test value and the median value of the normal range at time $t$ of a case *i*. $d'_i$ is the difference between the clinical laboratory test value and the median of the normal range of the clinical laboratory test at time *t+1* performed next to the time $t$ . Here, if the examination value at each time $t$ , *t+1* is within the normal value range, set the value of $d_i, d'_i$ to 0 to give discontinuity between the distance within the normal value range and the distance outside of the range. Thus, evaluation method 2 interprets a medication pattern with a smaller final evaluation

value as a pattern that can better improve the inspection value. For example, Figure 3 shows that the final evaluation value is calculated by the following formula:

$$\sqrt{\frac{d'_1+d'_2+0+d'_4}{4}}-\sqrt{\frac{d_1+d_2+d_3+d_4}{4}}\ (\because d'_3=0)$$

### 3.3 Procedure

Figure 4 shows the evaluation procedure. The procedure is outlined below.

1. Patient data for a specific condition for analysis are extracted from the EMRs based on the information from the current clinical laboratory test value state and the previous prescription medicine.
2. Prescription sequences are extracted by applying the proposed method and the existing method to the patient data; then, the next medicine to be prescribed is analyzed.
3. Regarding the next prescribed medicine extracted by each method, the evaluation value for the improvement of the clinical laboratory test value is calculated by the two evaluation methods described in Section 3.2. Then, the average value is compared with the top $k$ evaluation values.

We describe each of the above steps in detail. In step (1), patient data for a specific condition were first extracted based on clinical laboratory test value state patterns. The clinical laboratory tests we use are TSH (thyroid stimulating hormone), FT3 (free triiodothyronine) and FT4 (free thyroxine). In our experiment, we focus on the following three patterns of the clinical laboratory test value state:

- TSH, FT3, and FT4 are normal values.
- TSH is an abnormal value (Low); FT3 and FT4 are normal values.
- TSH is an abnormal value (High); FT3 and FT4 are normal values.

For each of these three patterns of the clinical laboratory test value state, we consider two previous prescription medicines patterns: thiamazole 5 mg and thiamazole 10 mg. The evaluation was conducted based on the combination of the patient's condition using six (3 × 2) patterns in total. In step (2), prescription sequence mining was performed for the six total patterns described above using the existing method based only on frequency information and the proposed method that considered unbiased frequent use by doctors. Then, the medicine to be prescribed next in the same situation was analyzed based on the prescription sequences extracted by each method. In step (3), the evaluation values for the improvement of the clinical laboratory test values were calculated for each prescription sequence extracted in step (2) using the two evaluation methods described in Section 3.2. Then, by comparing the average value of the top $k$ evaluation values of the prescription sequence extracted by each method, we examined how the proposed and existing methods properly extracted patterns that resulted in improvement of the clinical laboratory test values and evaluated the objective effectiveness of the extracted patterns by considering unbiased frequent use by doctors. Moreover, to measure the performance of top $k$ ranking of the existing and proposed method, we used a weighted average value based on the rank. This formula is defined as follows:

$$\frac{1}{k}\sum_{i=1}^{k}\frac{SCORE_i}{log_2(i+1)}$$

$SCORE_i$ shows the evaluation value for the improvement of the clinical laboratory test value in the medication pattern of ranking $i$ calculated by each evaluation method. For a value of $k$, we considered the eventual recommendation of the pattern to the doctor; because the maximum number of

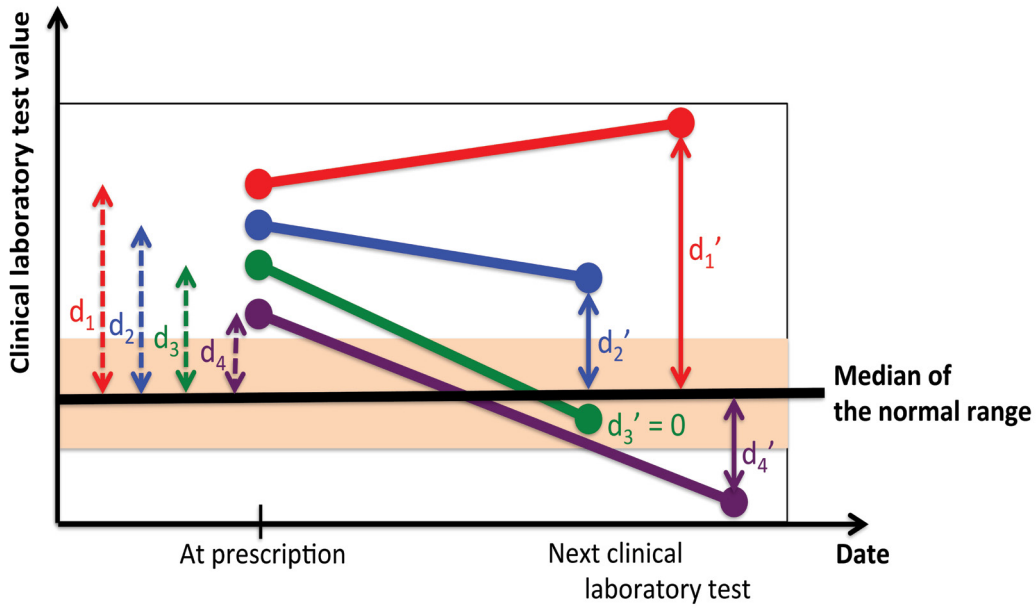



Figure 3: An example of a case group with an abnormal value (high).

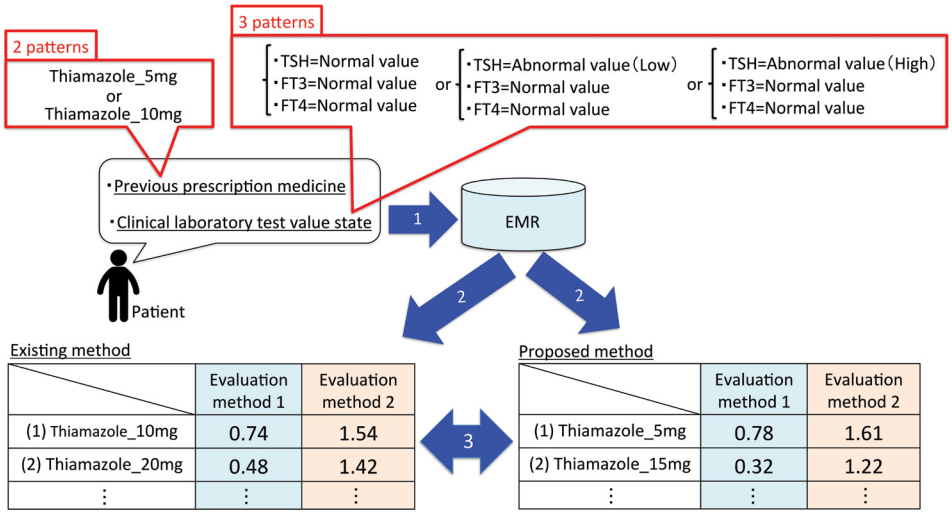



Figure 4: Evaluation procedure.

Table 1: Average values of the top $k$ (Evaluation method 1).

| | | | | | |
|---|---|---|---|---|---|
| **TSH, FT3, FT4=Normal/value** | | | | | |
| Existing method | | | | | |
| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
| Thiamazole_5 mg | 0.840 | 0.604 | 0.365 | 0.259 | 0.180 |
| Thiamazole_10 mg | 0.772 | 0.563 | 0.470 | 0.428 | 0.360 |
| Proposed method | | | | | |
| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
| Thiamazole_5 mg | 0.840 | 0.310 | 0.304 | 0.183 | 0.129 |
| Thiamazole_10 mg | 0.772 | 0.608 | 0.442 | 0.334 | 0.311 |
| **(TSH=Abnormal value(Low)) ∧ (FT3, FT4=Normal value)** | | | | | |
| Existing method | | | | | |
| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
| Thiamazole_5 mg | 0.232 | 0.173 | 0.159 | 0.162 | 0.143 |
| Thiamazole_10 mg | 0.231 | 0.080 | 0.080 | 0.057 | 0.060 |
| Proposed method | | | | | |
| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
| Thiamazole_5 mg | 0.397 | 0.280 | 0.234 | 0.195 | 0.174 |
| Thiamazole_10 mg | 0.189 | 0.086 | 0.095 | 0.059 | 0.058 |
| **(TSH=Abnormal value(High)) ∧ (FT3, FT4=Normal value)** | | | | | |
| Existing method | | | | | |
| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
| Thiamazole_5 mg | 0.612 | 0.428 | 0.367 | 0.343 | 0.312 |
| Thiamazole_10 mg | 0.536 | 0.426 | 0.386 | 0.342 | 0.305 |
| Proposed method | | | | | |
| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
| Thiamazole_5 mg | 0.612 | 0.494 | 0.446 | 0.381 | 0.335 |
| Thiamazole_10 mg | 0.501 | 0.419 | 0.349 | 0.327 | 0.299 |

Table 2: Average values of the top $k$ (Evaluation method 2).

| | | | | | |
|---|---|---|---|---|---|
| **TSH,FT3,FT4=Normal/value** | | | | | |
| Existing method | | | | | |
| | **$k=1$** | **$k=2$** | **$k=3$** | **$k=4$** | **$k=5$** |
| Thiamazole_5mg | 1.704 | 0.900 | 0.923 | 0.913 | 0.899 |
| Thiamazole_10mg | 1.602 | 1.349 | 1.494 | 1.286 | 1.170 |
| Proposed method | | | | | |
| | **$k=1$** | **$k=2$** | **$k=3$** | **$k=4$** | **$k=5$** |
| Thiamazole_5mg | 1.704 | 1.038 | 0.947 | 0.951 | 0.911 |
| Thiamazole_10mg | 1.602 | 1.286 | 1.161 | 1.077 | 1.138 |
| **(TSH=Abnormal value(Low)) ∧ (FT3,FT4=Normal value)** | | | | | |
| Existing method | | | | | |
| | **$k=1$** | **$k=2$** | **$k=3$** | **$k=4$** | **$k=5$** |
| Thiamazole_5mg | 0.219 | 0.172 | 0.160 | 0.138 | 0.123 |
| Thiamazole_10mg | -0.021 | 0.629 | 1.115 | 0.850 | 0.764 |
| Proposed method | | | | | |
| | **$k=1$** | **$k=2$** | **$k=3$** | **$k=4$** | **$k=5$** |
| Thiamazole_5mg | 0.171 | 0.171 | 0.148 | 0.129 | 0.120 |
| Thiamazole_10mg | 1.086 | 0.583 | 0.385 | 0.507 | 0.454 |
| **(TSH=Abnormal value(High)) ∧ (FT3,FT4=Normal value)** | | | | | |
| Existing method | | | | | |
| | **$k=1$** | **$k=2$** | **$k=3$** | **$k=4$** | **$k=5$** |
| Thiamazole_5mg | 6.755 | 4.570 | 2.633 | 1.864 | 1.481 |
| Thiamazole_10mg | 0.083 | 0.032 | 0.057 | 0.032 | 0.035 |
| Proposed method | | | | | |
| | **$k=1$** | **$k=2$** | **$k=3$** | **$k=4$T** | **$k=5$** |
| Thiamazole_5mg | 6.755 | 4.164 | 2.605 | 1.686 | 1.641 |
| Thiamazole_10mg | -0.029 | 0.011 | 0.041 | 0.026 | 0.013 |

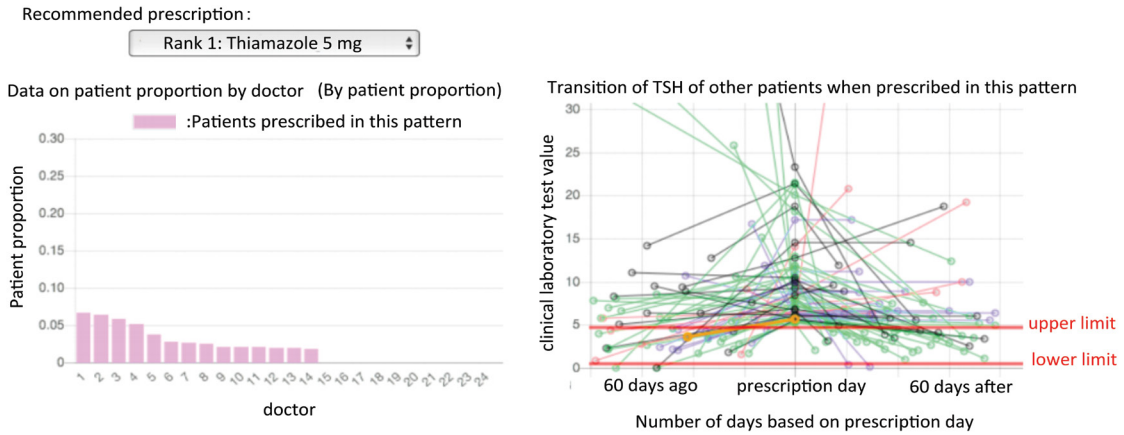



Figure 5: An example of user interface of predictions for the next medication for a disease [8].

elements that humans can recognize in short-term memory is five [7], we use $k$ values ranging from 1 to 5. In this experiment, the clinical laboratory test value item of the analysis target is TSH, because TSH reacts more sensitively even with a small amount of thyroid hormone excess or deficiency than FT3 and FT4 and can accurately capture small changes in the thyroid hormone.

## 4 Results

Table 1 and 2 show the performance result calculated by the formula (1) for the top $k$ ($k=1,2,\ldots,5$) using the evaluation method 1 and 2, respectively. Each table shows the performance results of the above-mentioned six total pattern combinations. The lower table is the results of the proposed method, and the upper table is the results of the existing method. Recall that larger the value the better the method in the evaluation method 1, and vice versa in the evaluation method 2.

## 5 Discussion and Conclusion

First, when TSH, FT3, and FT4 are within the normal value range, the existing method shows better performance in almost all cases, in the evaluation method 1. In the evaluation method 2, the existing method shows better performance when the previous prescription medicine was thiamazole (5 mg), while the proposed method gives better performance when the previous medicine was thiamazole (10 mg). Under these circumstances, we suggest that better improvement of the clinical laboratory test value can be expected in the existing method than the proposed method when TSH, FT3, and FT4 are within the normal value range.

Second, when TSH has an abnormal value (Low) and the FT3 and FT4 values are in the normal range, the proposed method has a better evaluation value in almost all cases in both evaluation methods 1 and 2. Thus, improvement of the clinical laboratory test value can be expected in the proposed method in this case.

Third, when TSH has an abnormal value (High) and the FT3 and FT4 values are in the normal range in evaluation method 1, the evaluation value of the existing method is better when the previous prescription medicine is thiamazole (10 mg), and the evaluation value of the proposed method is better when the previous prescription medicine is thiamazole (5 mg). In evaluation method 2, the proposed method has a better evaluation value in almost all cases. Thus, improvement of the clinical laboratory test value can be expected in the proposed method in this case.

In conclusion, the present study suggested that our prescription mining method is useful in certain situations such as when patients are not in normal condition (i.e. the clinical laboratory test value is outside of the normal value range). This result suggests our method supports medial doctors more efficiently to decide a reasonable choice for clinical improvement especially when such decision is difficult.

Finally, we discuss possible future extension of our work. In this study, the data mining methods used only prescription data and clinical laboratory test value data. However, various other information recorded in the EMRs including personal information, such as age, gender, weight, and history of disorder, can be effectively used for data mining. Progress notes written in natural language, various medical imaging data, and information from outside of the hospital (i.e., a life log) can also be effectively used in combination with our study. Other possible extension includes evaluation experiments using data other than thyroid diseases and data from multiple medical institutions. A prototype of user interface for suggestions of medications to prescribe next is shown in Figure 5 [8].

## 6 Acknowledgement

This work was partially supported by The Japan Society for the Promotion of Science (JSPS) KAKENHI grant number 15H02705 and 15K00466. The funding body had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 7 Authors' Contributions

M. Yoshikawa, A. Hamasaki, O. Sugiyama, K. Okamoto, and T. Kuroda have contributed to the conception and design of research; T. Kuroda, K. Okamoto and Y. Morita collected and managed the research data. Y. Morita has drafted the article and performed experiments. N. Kada developed a user interface of the system. M. Yoshikawa, A. Hamasaki, O. Sugiyama, K. Okamoto, and T. Kuroda performed the interpretation the data and results. M. Yoshikawa revised the article critically. All authors have given approval of the final version to be published.

## 8 Conflict of Interest

The authors declare no conflicts of interest associated with this manuscript.

## References

[1] http: //www.openclinical.org/

[2] Iguchi R, Sato H, Nakamura K, Matsubara T, Gunshin M, Ishii T, et al. Electronic medical system in emergency outpatient clinic in the united states and clinical decision support system. J National Institute of Public Health. 2013; 62: 88-97 (Japanese).

[3] Asai H, Jimbo M, Nease DE Jr., Ishibashi Y, Fetters MD. Using computerized clinical decision support systems for quality improvement of preventive and chronic care. An Official Journal of the Japan Primary Care Association. 2011; 34: 133-140 (Japanese).

[4] Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. J Biomed Inform. 2015; 53: 73-80.

[5] Hirano S, Tsumoto S. Clustering of order sequences based on the typicalness index for finding clinical pathway candidates. In: Ding W, Washio T, Xiong H, Karypis G, Thuraisingham B, Cook D, et al. editors. ICDMW 2013: Proceedings of IEEE the 13th International Conference on Data Mining Workshops; 2013 Dec 7-10; Dallas, USA. Los Alamitos: IEEE; 2013.

[6] Mohammed JZ. SPADE: An efficient algorithm for mining frequent sequences. Machine learning. 2001; 42: 31-60.

[7] Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behav Brain Sci. 2001; 24: 87-114.

[8] Kada N, Yoshikawa M, Hamasaki A, Sugiyama O, Okamoto K, Kuroda T. User interface for clinical decision support system in medication for thyroid disease. DEIM Forum 2017, 2017 (Japanese).

# Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days

**Kazuya Okamoto[1,2]*, Toshio Uchiyama[3#], Tadamasa Takemura[4], Naoto Kume[2], Tomohiro Kuroda[1,4] and Hiroyuki Yoshihara[2]**

[1] Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan

[2] Graduate School of Medicine, Kyoto University, Kyoto, Japan

[3] Faculty of Business Administration and Information Science, Hokkaido Information University, Hokkaido, Japan

[4] Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan

## Abstract

**Objectives:** The authors developed and evaluated a method of selecting accurate diagnosis procedure combination (DPC) codes based on standardized treatment information relative to the number of hospitalization days.

**Methods:** The authors used machine learning methods to generate DPC codes based on treatment data. The machine learning methods utilized were the Naïve Bayes method, the SVM method, and a combined method of the two methods. We prepared DPC code data and standardized treatment data corresponding to cases occurring in fiscal year 2008 at Kyoto University Hospital. To produce classification models, machine learning methods require a moderate amount of data corresponding to each DPC code; accordingly, we selected 166 DPC codes that were each related to at least 20 cases. The number of cases with these DPC codes was 10,123.

**Results:** DPC code selection was attempted using the Naïve Bayes method, the SVM method, and the combined method of the two; of these, the combined method yielded the best results, producing accurate DPC codes in 73.8% of cases. In addition, we were able to improve the precision of DPC code selection to 76.5% by utilizing partial treatment data gathered up to the 11th day of each hospitalization.

**Conclusion:** The present study confirmed the feasibility of automatic DPC code selection through machine learning methods based on treatment information. Our future work will include the construction of a system to select DPC codes automatically and the evaluation of this system to determine whether it can reduce doctors' workloads.

## Keywords

DPC; Treatment information; Artificial intelligence; Machine learning

**Correspondence to:**

**Kazuya Okamoto**
Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan.
E-mail: kazuya@kuhp.kyoto-u.ac.jp
#This work was performed when T. Uchiyama was affiliated with NTT Corporation.

**EJBI 2018; 14(1):45-51**
Received: December 15, 2017
Accepted: January 29, 2018
Published: February 05, 2018

## 1 Introduction

The diagnosis procedure combination (DPC) is a diagnosis group classification created in imitation of the diagnosis-related groups (DRGs) used in the US. DRGs and DPC are classifications according to principal diagnosis, the presence of complications, operations and treatments, and other relevant criteria. DRGs and DPC have been used to analyze diseases and hospital managing statuses [1, 2, 3, 4]. Thomas and Ashcraft have showed DRGs to be highly reliable to measure the severity of illness of patients [1]. Stang, Merrill and Kuss have calculated population-wide rates of hysterectomy across Germany on the basis of nationwide DRG statistics [2]. Lotter et al. have compared reimbursement of burns based on DRGs in four European countries [3]. Nakagawa, Yoshihara and Nakagawa have made indicators to evaluate the management efficiency and

medical activity in each DPC or DRG group [4]. In 2003, Japan introduced a payment system of medical fees based on DPC. Like the DRG-based payment system used in the US, the DPC-based payment system is a prospective payment system and is known as the DPC/PDPS (Per-Diem Payment System). Essentially, the medical fee associated with any particular hospitalization is determined based on the DPC assigned to that hospitalization. In hospitals that have adopted the DPC/PDPS system, doctors are requested to assign one DPC code to each case in order to calculate the medical fee associated with that case. In fiscal year 2011, the DPC/PDPS system was used in 1,449 hospitals and the beds in these 1,449 hospitals made up almost half of all hospital beds in Japan.

Selecting a DPC code for a case virtually never directly improves medical care, and the task regularly takes up time that could otherwise be devoted to patient care. Consequently, in some divisions of Kyoto University Hospital, chief physicians or assistant chief physicians select the DPC codes for all hospitalizations, taking the burdensome selection process upon them to save doctors' time. Additionally, Suzuki et al. have pointed out the difficulty of accurately selecting only one DPC code for each patient because many patients have more than one disease [5]. For this reason, Suzuki et al. examined the capability of machine learning methods to select suitable DPC codes automatically based on the contents of discharge summaries [5]. Similarly, Okamoto et al. have improved the precision of automatic DPC code selection based on discharge summaries by testing several machine learning methods and combined methods of the machine learning methods [6]. They showed that DPC codes can be accurately assigned to cases automatically. However, the use of data from discharge summaries for this purpose will not help reduce doctors' workloads because the discharge summary is usually written after the patient's discharge and the selection of a DPC code is required for the patient to be discharged, meaning that a DPC code must be selected before the discharge summary is written.

Therefore, in order to reduce doctors' workloads, the authors attempt to generate a means of accurate automatic DPC code selection based on treatment information which is stored before patient discharge.

## 2 Objectives

In this paper, the authors develop a method of accurate automatic DPC code selection based on standardized treatment information and evaluate the developed method. In the course of hospitalization, the patient undergoes a series of treatments, and thus, treatments are different depending on the phases of the hospitalization. Moreover, doctors are requested to decide a DPC code in an early date of hospitalization. Therefore, we evaluate how precisely DPC codes are selected based on partial treatment data relative to the number of hospitalization days.

## 3 Materials and Methods

### 3.1 DPC Codes

Hospitals adopting the DPC/PDPS system are required to submit DPC codes for each admission case to the Japanese government, which is a part of data set called File Format 1. DPC codes consist of 14 digits. The first six digits of a DPC code indicate the disease name while the remaining eight digits indicate disease severity, complications, the age of the patient, the purpose of admission, operations required, treatments needed, and other variables. For instance, in the DPC code 060030xx01x0xx, the first six digits (060030) indicate that the disease was malignant tumors of the small bowel and the remaining eight digits (xx01x0xx) show that the patient underwent small bowel resection. In fiscal year 2008, the number of DPC codes used in the DPC/PDPS system was 1,572 and the number of diseases indicated by the 1,572 DPC codes was 452. In fiscal year 2011, the number of DPC codes used was 1,880 and the number of diseases indicated by the 1,880 DPC codes was 474.

### 3.2 Treatment Data

Hospitals adopting the DPC/PDPS system must submit data on medical treatments to the Japanese government. The data on medical treatments are standardized and sorted into three types of files known as File D, File E and File F. Files E and F include complete treatment information; File E includes medical billing information according to the medical treatment fee system; File F includes more detailed information according to the daily medical records which corresponds to the minimum unit of the master file of medical materials or examinations, etc. File D, in contrast, includes only treatment information which is not included in terms of payment in the DPC/PDPS system. For instance, microscopic examination of pathological tissue is a treatment which is included in terms of payment in DPC/PDPS system, and so File D does not include the information of this; Pathologic diagnosis expense is a treatment which is not included in terms of payment in DPC/PDPS system, and so File D includes the information of this.

### 3.3 Machine Learning Methods

In this study, the authors use machine learning methods known as supervised learning, which learn a classified training set and then classify a sample set. These methods thus require a classified training set.

We describe the Naïve Bayes method [7] and the SVM (support vector machine) method [8], both of which are useful machine learning methods. Several previous studies have shown that the precision of automatic classification can

increase when machine learning methods are combined [9, 10, 11, 12]. In particular, Uchiyama et al. have shown that a combined method of machine learning methods using corresponding levels of confidence calculated by logistic regression analysis is effective [12]. Okamoto et al. have likewise demonstrated that the combined methods are effective means of selecting DPC codes based on discharge summaries [6]. Hence we explain not only the Naïve Bayes method and the SVM method but also their combined method of machine learning methods.

**Naïve bayes method**

The Naïve Bayes method measures frequencies of elements in a training set to obtain a classification model [7]. In this paper, we utilize a multinomial model, which takes into consideration occurrence frequencies of elements in data. For data in a sample set, the multinomial model calculates the likelihood of each class and classifies data into the class whose likelihood is highest in all classes. Note that, when we count the number of elements, we smooth the data using the Good-Turing method to prevent zero probability [13].

**SVM**

A support vector machine (SVM) [8] finds a hyper plane which can effectively separate a training set described in a vector space into two classes. Thus SVMs are basically binary classifiers. Here, we use linear SVMs, which are the simplest SVMs. Linear SVMs are known to be effective when the number of vector elements is large, as it often is in document classification. When SVMs are applied to multiclass classification problems, they must be combined and a class must be selected based on proper criteria. In this study, we use the one-versus-the-rest classification [14], the most popular SVM technique.

**Combined method**

In this study, we also utilize a combined method of machine learning methods that was proposed by Uchiyama et al. [12]. First, the combined method eliminates levels of confidence of machine learning methods through logistic regression analysis. Confidence is an expression of the correctness of each machine learning method. When the combined method eliminates corresponding levels of confidence, it uses *classification scores* of the most probable class and the second most probable class. Classification scores are indicators based on which machine learning methods classify data, for example, likelihood in the Naïve Bayes method and distances from hyperplanes in SVMs. A level of confidence $P$ is calculated as follows:

$$P = \frac{\exp Z}{1 + \exp Z} \tag{3}$$

$$Z = \beta_0 + \beta_1 L_1 + \beta_2 L_2 \tag{4}$$

where $\beta_0$, $\beta_1$, $\beta_2$ are model parameters and $L_1$, $L_2$ are classification scores of the most probable class and the second most probable class. The combined method calculates the values of model parameters $\beta_0$, $\beta_1$, $\beta_2$ by using logistic regression analysis corresponding to every machine learning method and every class in a class set.

Next, the combined method classifies sample data through the following steps: i) It classifies the sample data using more than one machine learning method. ii) It calculates corresponding levels of confidence based on classification scores of the most probable class and the second most probable class corresponding to every machine learning method. iii) It classifies the sample data into the class selected by the machine learning method with the highest level of confidence.

## 3.4 Experiments

The authors used machine learning methods to accurately select DPC codes based on treatment data. Namely, the DPC code data were output variables of the machine learning methods and all treatment data included in Files D, E and F were input variables of the machine learning methods.

We prepared data corresponding to cases occurring in fiscal year 2008 at Kyoto University Hospital. In the cases, every patient was admitted after April 1st 2008 and was discharged before March 31st 2009. The number of cases submitted to the DPC/PDPS system was 14,460 and the number of DPC codes related to these cases was 1,133. When generating classification models, machine learning methods require a moderate amount of data corresponding to each class, i.e., each DPC code, and thus, we only chose DPC codes that were each assigned to at least 20 cases. The number of chosen DPC codes was 166 and the number of cases assigned to these DPC codes was 10,123. Table 1 shows the top-10 list of DPC codes in terms of the number of cases.

The machine learning methods we used in experiments were the Naïve Bayes method, the SVM method and the combined method of the two methods (introduced in 3.3.3). Note that we implemented the Naïve Bayes method, used SVM-Light [15] utilizing the SVM method, and created the combined method. In the Naïve Bayes method, elements and the occurrence frequencies of the elements correspond to treatment types and the number of treatments, respectively. In the SVM method, likewise, elements of vectors and the values of the elements correspond to treatment types and the number of treatments, respectively.

**Experiment 1**

We evaluated the feasibility of accurate automatic selection of DPC codes by applying the Naïve Bayes method, the SVM method and the combined method of the two methods to the

Table 1: Target DPC codes and the number of the cases (top 10).

| DPC code | DPC name | # of cases |
|---|---|---|
| 020200xx970xxx | Macula and posterior pole degeneration, presence of an operation. | 352 |
| 060050xx97100x | Cancer of liver and intrahepatic bile duct, presence of another operation, presence of operation and treatment 1. | 291 |
| 050050xx9910xx | Angina and chronic ischemic heart disease, presence of operation and treatment 1. | 242 |
| 020110xx97x0x0 | Cataract and disease of the lens, presence of an operation, one eye. | 215 |
| 100070xxxxxx0x | Type 2 diabetes except for diabetic ketoacidosis. | 201 |
| 070560xx99x0xx | Autoimmune disease with generalized organ dysfunction. | 198 |
| 020110xx97x0x1 | Cataract and disease of the lens, presence of an operation, both eyes. | 194 |
| 120020xx99x41x | Cancer of cervix and uterine body, presence of operation and treatment 2.4, presence of a complication. | 180 |
| 030250xx991xxx | Sleep apnea, presence of operation and treatment 1. | 179 |
| 120010xx99x31x | Cancer of ovary and uterine adnexa, presence of operation and treatment 2.3, presence of a complication. | 174 |

chosen cases. The evaluation was performed via 10-fold cross-validation. We measured the precision of the 14 digits of the automatically selected DPC codes and the precision of their first six digits, which indicate the disease name. Furthermore, because an automatic DPC code selection system that proposes several candidate DPC codes could be effective, we had the methods generate five possible 14-digit DPC codes and measured their ability to produce one correct DPC code out of the five. We also had the methods generate five possible candidates of the first six digits of the DPC codes and measured their ability to produce one correct six-digit sequence. These four measurements, namely, the precision of the 14 digits of a single automatically selected DPC code, the precision of the first six digits of a single automatically selected DPC code, the precision of one of five suggested 14-digit DPC codes, and the precision of one of five suggested first-six-digit sequences, are referred to as *the 14-digit top-1 precision*, *the 6-digit top-1 precision*, *the 14-digit top-5 precision* and *the 6-digit top-5 precision*, respectively. However, the combined method cannot be used to select additional DPC codes, because a combined method of machine learning methods used here selects a single one DPC code by selecting one machine learning method. Hence, the combined method selects the remaining four of the five possible DPC codes according to the Naïve Bayes method.

**Experiment 2**

We evaluated how precisely DPC codes are selected based on partial treatment data relative to the number of hospitalization days. Each machine learning method makes a classification model by using treatment data up to the $x$-th day of hospitalization and classifies a sample set consisting of treatment data up to the $x$-th day of hospitalization by utilizing the classification model. In this experiment, we evaluated the precision achieved using the Naïve Bayes method, the SVM method and the combined method of the two methods when $x = 1, \ldots, 20$. As in experiment 1, we measured the 14-digit top-1 precision, the 6-digit top-1 precision, the 14-digit top-5 precision and the 6-digit top-5 precision. Note that, in order to choose five candidates, the combined method selects the remaining four DPC codes according to the Naïve Bayes method, as in experiment 1.

Table 2: The results of automatic DPC code selections based on treatment data using the Naïve Bayes method, the SVM method and the combined method of the Naïve Bayes method and the SVM.

| Machine Learning Methods | Precision (%) | | Prec. of top 5 (%) | |
|---|---|---|---|---|
| | 14 digits | 6 digits | 14 digits | 14 digits |
| Naïve Bayes method | 63.2 | 71.5 | 87.6 | 90.6 |
| SVM | 72.1 | 75.7 | 86.4 | 89.8 |
| Combined method | 73.8 | 78.7 | 89.9 | 93.2 |

## 4 Results

The results of experiment 1 are shown in Table 2. The second and third columns show the 14-digit top-1 precision and the 6-digit top-1 precision, respectively. The fourth and fifth columns show the 14-digit top-5 precision and the 6-digit top-5 precision, respectively.

Figures 1-4 show the 14-digit top-1 precision, the 6-digit top-1 precision, the 14-digit top-5 precision and the 6-digit top-5 precision, respectively, in experiment 2. In these figures, the Naïve Bayes method and the combined method of the Naïve Bayes method and the SVM method are abbreviated as NB and NB+SVM, respectively. Moreover, NB+SVM (all)

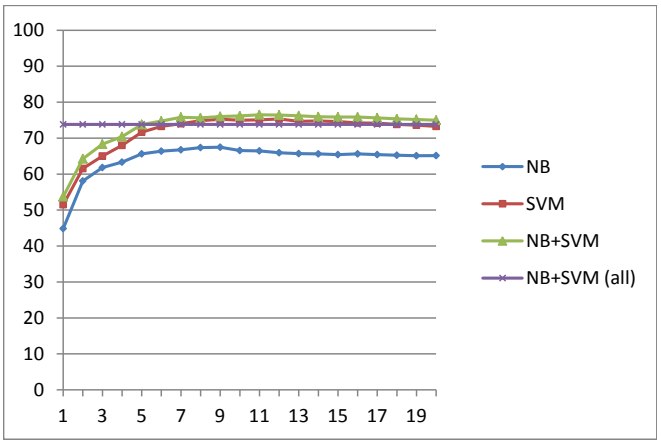



Figure 1: The precision of the 14 digits of automatically selected DPC codes relative to the number of hospitalization days from which data was used (the 14-digit top-1 precision).

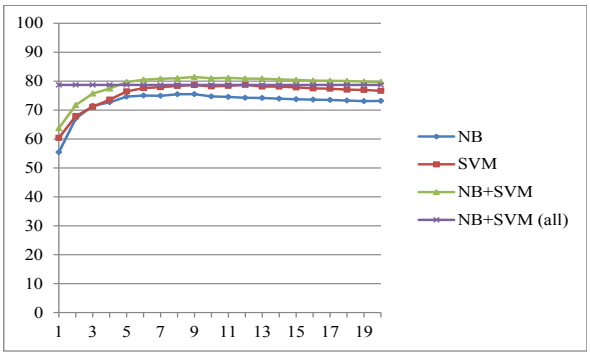



Figure 2: The precision of the first six digits of automatically selected DPC codes relative to the number of hospitalization days from which data was used (the 6-digit top-1 precision).

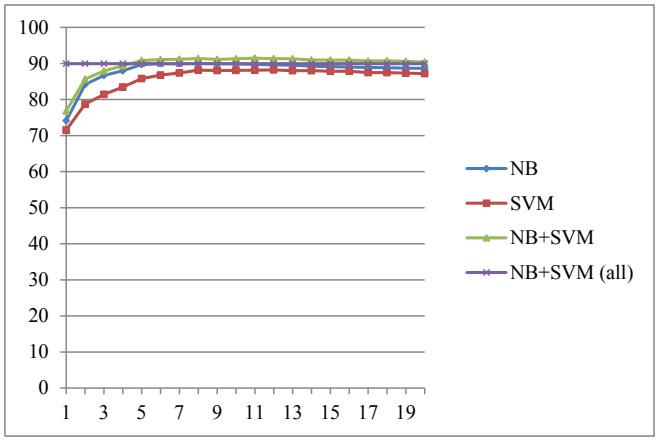



Figure 3: The precision of one of five suggested 14-digit DPC codes relative to the number of hospitalization days from which data was used (the 14-digit top-5 precision).

indicates the combined method which makes a classification model by using the treatment data of entire hospitalization as a training set and selects DPC codes by using a sample set consisting of the treatment data of entire hospitalization. The results of NB+SVM (all) are shown for comparison.

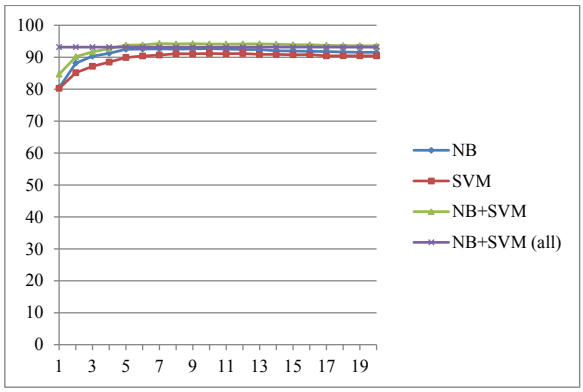



Figure 4: The precision of one of five suggested first-six-digit sequences relative to the number of hospitalization days from which data was used (the 6-digit top-5 precision).

## 5 Discussion

In experiment 1, the combined method of the Naïve Bayes method and the SVM method yielded results that were more precise than those of either the Naïve Bayes method or the SVM method. The 14-digit top-5 precision and the 6-digit top-5 precision achieved using the combined method were almost 90% and over 90%, respectively; these levels of precision demonstrate that the combined method is useful. In experiment 2, all four precision values were maximized as the number of hospitalization days from which data were used approached 10, showing that DPC code selection based on partial treatment data relative to the number of hospitalization days is ideal and that characteristic treatments are tend to be administered at the beginning of hospitalization. In actuality, characteristic treatments, such as operations and expensive medicines, are administered at the beginning of hospitalization and a DPC code seems to be determined based on the characteristic treatments.

Horiguchi et al. proposed several rules by which to determine the DPC codes for cases to which disease names had already been assigned, and showed that 95.8% of cases were assigned proper DPC codes according to their rules [16]. In particular, they pointed out that when the first six digits of a DPC code are given, the remaining eight digits can be automatically selected with high precision. However, since they did not clearly divide their data into a training set and a sample set, their study did not have sufficient cross-validation. In the present study, the 14-digit top-1 precision of the combined method was 73.8% and the 6-digit top-1 precision of the combined method was 78.7%. The results confirmed that when the first six digits of a DPC code are given, the remaining eight digits can be automatically selected with high precision of 93.8%.

Nishi and Uozawa selected characteristic treatment data from Files E and F and constructed a decision tree using this data which was able to correctly determine the first six digits

of the DPC code in 74.5% of cases [17]. Since they selected the treatment data which they considered characteristic, however, the data were not standardized. In the present study, in experiment 1, the combined method yielded the first six digits of the appropriate DPC code with a higher level of precision: 78.7%. In experiment 2, the precision increased to 81.5% when data up to the 9th day of hospitalization were incorporated. Although it is difficult to compare the results of the present study with those of Nishi and Uozawa because of difference data sets used in each study, the precision of our results was higher in spite of the smaller size of training data and the larger number of possible DPC codes. The results achieved in the present study appear to be superior to those achieved in the study by Nishi and Uozawa.

In experiments 1 and 2, when the combined method selected five candidates, it preferred the Naïve Bayes method. If the combined method were to prefer the SVM method, the precision could increase. This is not likely, however, given that, in the other experiments where the combined method preferred the SVM method, the 14-digit top-5 precision and the 6-digit top-5 precision were 87.7% and 91.1%, respectively, which were lower than the levels achieved when the combined method preferred the Naïve Bayes method. Additionally, in DPC code selections using partial treatment data relative to the number of hospitalization days, the precision achieved when the combined method preferred the Naïve Bayes method was always higher than the precision achieved when the combined method preferred the SVM method.

In experiment 2, the machine learning methods made classification models by using partial treatment data from up to the $x$-th day of hospitalization to classify a sample set consisting of partial treatment data from up to the $x$-th day by utilizing the classification models. We compared this technique with the machine learning methods using the treatment data from the entire hospitalization as a training set. As $x$ increased, the precision of the alternative methods approached asymptotically to the precision of the machine learning methods trained on treatment data from the entire hospitalization to select DPC codes using a sample set consisting of treatment data from the entire hospitalization. Also, the precision of the alternative methods was always less than the precision of the machine learning methods used in experiment 2 when $x = 1, \ldots, 20$. Thus we confirmed that a patient with a particular diagnosis does tend to undergo a series of treatments and that machine learning methods can learn effectively when trained on partial treatment data relative to the number of hospitalization days.

We consulted chief physicians in some divisions about the implications of automated coding in practical terms. According to them, they take a few minutes to select each DPC code and feel the processes laborious especially when they must change DPC codes. Though their comments cannot always reflect practical situations, we obtained suggestions about effective use of an automated coding system. If automatic selection of DPC codes by machine learning methods is to be implemented to reduce doctors' workloads, one problem will have to be resolved. Machine learning methods need a moderate amount of data corresponding to each class. In the present study, we selected 166 DPC codes; in contrast, the number of DPC codes used in the DPC/PDPS system in fiscal year 2008 was 1,572. However, the 166 DPC codes used in the present study were assigned to 70.0% of the actual cases submitted to the DPC/PDPS system. To confirm the viability of automatic DPC code selection using machine learning methods, we consulted a doctor, who pointed out that, even if machine learning methods could be used to identify only partial DPC codes for only 70.0% of cases that alone would significantly reduce doctors' workloads. Another drawback is that the DPC codes used in the DPC/PDPS system are revised every two years. Nevertheless, it is not difficult to translate the codes of an old DPC code system into a new version.

## 6 Conclusion

The authors evaluated automatic DPC code selection by machine learning methods based on treatment information for the purpose of reducing doctors' workloads. In 73.8% of cases, accurate DPC codes were selected when the combined method of the Naïve Bayes method and the SVM method was used. In addition, we were able to improve the precision of DPC code selection to 76.5% by incorporating partial treatment information from up to the 11th day of hospitalization. Our future work will include the construction of a system that selects DPC codes automatically and the evaluation of this system to determine whether it can reduce doctors' workloads.

## 7 Acknowledgement

This work was supported by KAKENHI (15K00466).

## 8 Author Contribution

K. Okamoto, T. Uchiyama, T. Takemura and H. Yoshihara have contributed to the conception and design of research; K. Okamoto, T. Takemura, N. Kume and T. Kuroda collected and managed the research data; K. Okamoto and T. Uchiyama performed experiments; K. Okamoto, T. Uchiyama, and T. Takemura performed the interpretation the data and results. K. Okamoto and T. Uchiyama has drafted the article; K. Okamoto, T. Uchiyama and T. Kuroda revised the article. All authors have given approval of the final version to be published.

## 9 Conflict of Intrest

The authors declare no conflicts of interest associated with this manuscript.

## References

1. Thomas JW, Ashcraft ML. Measuring Severity of Illness: a Comparison of Interrater Reliability among Severity Methodologies. Inquiry. 1989; 26: 483-492.

2. Stang A, Merrill RM, Kuss O. Hysterectomy in Germany: a DRG-based nationwide analysis, 2005-2006. Dtsch Arztebl Int. 2011; 108: 508-514.

3. Lotter O, Jaminet P, Amr A, Chiarello P, Schaller HE, Rahmanian-Schwarz A. Reimbursement of burns by DRG in four European countries: An analysis. Burns. 2011; 37: 1109-1116.

4. Nakagawa Y, Yoshihara H, Nakagawa Y. New Indicators Based on Personnel Cost for Management Efficiency in a Hospital. J Med Syst. 2011; 35: 625-637.

5. Suzuki T, Yokoi H, Fujita S, Takabayashi K. Automatic DPC Code Selection from Electronic Medical Records: Text Mining Trial of Discharge Summary. Methods Inf Med. 2008; 47: 541-548.

6. Okamoto K, Uchiyama T, Takemura T, Adachi T, Kume N, Kuroda T, et al. Automatic DPC Code Selection from Discharge Summaries Using Several Machine Learning Methods. Trans Japanese Soc Med Bio Eng. 2011; 49: 40-47 (Japanese).

7. McCallum A, Nigam K. Comparison of Event Models for Naïve Bayes Text Classification. Proceedings of AAAI-98 Workshop on Learning for Text Categorization; 1998. p. 41-48.

8. Boser BE, Guyon IM, Vapnik VN. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory; 1992. p. 144-152.

9. Schapire RE, Singer Y. Improved Boosting Algorithm Using Confidence-rated Predictions. Machine Learning. 1999; 37: 297-336.

10. Schapire RE, Singer Y. BoosTexter: A Boosting-based System for Text Categorization. Machine Learning. 2000; 39: 135-168.

11. Kudo T, Matsumoto Y. A Boosting Algorithm for Classification of Semi-Structured Text. Trans IPS Japan. 2004; 45: 2146-2156 (Japanese).

12. Uchiyama T, Bessho K, Uchiyama T, Oku M. A Multi-classifier Combination Using Confidence Estimation. Proceedings of the 84th Knowledge Base Systems; 2009. p. 33-38 (Japanese).

13. Chen SF, Goodman J. An Empirical Study of Smoothing Techniques for Language Modeling. Technical Report at Harvard University. 1998; TR-10-98.

14. Hamel LH. Knowledge Discovery with Support Vector Machines. Massachusetts: Wiley-InterScience; 2009.

15. Joachims T. Making large-Scale SVM Learning Practical. In: Schölkoph B, Burges CJC, Smola AJ, editors. Advances in Kernel Methods: Support Vector Learning. Massachusetts: MIT-Press; 1999.

16. Horiguchi H, Yasunaga H, Hashimoto H, Ishikawa KB, Kuwabara K, Anan M, et al. Development of the Standard DPC Coding Logic. J Med Inf 2008; 28: 73-82 (Japanese).

17. Nishi T, Uozawa M. The DPC Disease Decision Model from Medical Treatment Data by Data Mining. In: Proceedings of the 29th Joint Conference on Medical Informatics; 2009: p. 56-459 (Japanese).

# Change Detection in Vital Signs Associated with Impending Death for Homecare Patients Using a Pressure-Sensing Mat

**Takayuki Shuku[1]*, Noriko Sakano[2], Mizuki Morita[2] and Shingo Kasahara[2]**

[1] Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan

[2] Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama University, Okayama, Japan

## Abstract

How vital signs change at the end of life is a frequently asked question by health-care professionals, patients, and families, and research groups have attempted to identify specific changes and indicators associated with impending death. This study proposed a new method for detecting characteristic changes in vital signs associated with patient's death based on time-series data measured by a pressure-sensing mat. The pressure-sensing mat is less invasive and can automatically log heart and respiratory rates with high temporal resolutions. In the proposed method, the probability density functions (PDFs) of vital signs are estimated based on the large amount of data, and temporal changes in the PDFs are evaluated using an information measure, the Jenson-Shannon divergence (JSD). The proposed method was used to detect changes in vital signs of homecare patients to demonstrate its applicability. By tracking the temporal changes in the PDFs and JSD of vital signs, the specific changes associated with death can be detected. The proposed method can be a promising approach in quantitative and simplified prediction of impending death in homecare patients.

## Keywords

Vital signs; Pressure-sensing mat; Change detection; Probability density functions

**Correspondence to:**

**Takayuki Shuku**
Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan.
E-mail: shuku@cc.okayama-u.ac.jp

**EJBI 2018; 14(1):52-57**
Received: December 28, 2017
Accepted: February 13, 2018
Published: February 20, 2018

## 1 Introduction

Clinicians observe decline in physical function (or activities of daily living, ADLs) in patients at the end of their lives, and patients experience a multitude of physiological changes such as muscle weakness, dysphagia, and altered level of consciousness [1, 2]. In addition to these physiological changes, vital signs such as heart rate, blood pressure, and respiratory rate also provide useful information on the patient's health status, and these changes can signify impending death [3]. Accurate prediction of impending death is of great importance to clinicians who attend to patients at the end of life because it could affect their communication with patients and families and inform complex decision making such as discontinuation of investigations and aggressive treatments, discharge planning, and enrollment into clinical care pathways.

Several studies have examined diagnostic signs of impending death. Bruera et al. [4] investigated variation in vital signs in the last days of life in patients and reported that blood pressure and oxygen saturation decreased as patients approach death. Hui et al. [5] identified eight highly specific physical signs associated with death within 3 days among cancer patients. The published studies, however, were designed for hospitalized patients with advanced cancer, and the applicability of their findings to homecare patients is extremely limited. Development of simplified methodologies to detect the changes associated with impending death for homecare or non-cancer patients is necessary. Some research groups have developed methods of predicting death in homecare patients, and their predictions have performed well [6, 7, 8]. Many existing methods, however, were developed to predict relatively long survival times (e.g., up to 10 years). Long-term predictions have limited application in immediate care planning and decision making, especially for homecare patients, who may be in their final days of life, and their families. In addition, most existing prognostic methods are not adaptive and were not developed as web applications.

Recently, sophisticated sensors designed to measure vital signs have been developed, and a pressure-sensing mat has been paid great attention because it is less invasive and can automatically measure heart and respiratory rates with high temporal resolution. It would be a promising approach to identify characteristic changes associated with impending death based on large amounts of data measured by the pressure-sensing mats. A number of studies have developed less/non-invasive monitoring using pressure sensors [9, 10, 11]. These sensors are basically designed to evaluate sleep quality or to automatically detect sleep-wake states, but they have never used for predicting impending death for patients [12] outlined how future health care is shaping up current research efforts in devices and sensors, signal processing, and machine learning and reported the potential of sensor data-driven prognosis.

This study proposes a new methodology to detect the changes in vital signs associated with impending death based on the high-temporal-resolution data measured by a pressure-sensing mat and to analyze the data using a machine learning method, i.e., density ratio estimation [13]. This paper describes the concept of the method proposed in this study and the application example for actual vital signs measured by a pressure-sensing mat.

## 2 Materials and Methods

### 2.1 Sensor Mat

We used the sensor mat Odayaka Time developed by FSC Co., Ltd. (Nagoya, Japan) to measure time-series vital signs for homecare patients (shown below). This sensing mat can be placed on the top of any mattress and measure pressure based on the patient's body motion in bed. Then, vital signs are estimated based on the measured pressure. The mat continuously takes measurements and is flexible, thin, and unnoticeable.

The sensor does not properly work when the patient's back is off the mat and sometimes cannot measure data. We defined the missing data rate and used it as an auxiliary data for detecting changes in patient's status.


Pressure-sensing mat: Odayaka Time.

### 2.2 Data Collection

Although 180 patients have been equipped with this mat, we only collected time-series data from two bedridden homecare patients, patients 1 and 2, whose vital signs were measured until death. This is because the other data are not consistent and appropriate for the analysis. Information about the two patients is summarized in Table 1. The measurement period for patients 1 and 2 was 8 and 3 months, respectively. The patient requirement and study design were approved by the ethics committee of Okayama University, and informed consent was obtained from the patients or their families.

### 2.3 Data Analysis based on Density Ratio Estimation

We developed a new method for detecting changes in time-series vital signs using density ratio estimation [13]. The methods for change/anomaly detections in time-series data include dynamic time-warping method (TWM) [14], singular spectrum transformation (SST) [15], and autoregressive (AR) model-based method [16]. These methods, however, have some drawbacks in practical applications, such as sensitivity to noise (TWM), time consuming (SST), and inability to detect drastic changes in time-series data (AR model-based method). Recently, the more advanced method of change detection, called density ratio estimation, has received much attention because of its simplicity and robustness and its application in time-series data analysis [13].

To develop the new method, we made the following assumptions:

1. The value of the vital sign is a sample generated from the inherent probability density functions (PDFs) of the patients.
2. Change in the PDFs shows the patient's health status.
3. Specific changes in vital signs associated with impending death can be detected by determining the temporal change in the PDFs.

The PDFs of the heart and respiratory rates are estimated based on the time-series data measured by the pressure-sensing mat. The histogram of the vital signs was drawn using the data in specific time intervals (24 hours in this study). Since the temporal resolution of the data is 1 minute, one PDF is represented by 1,440 discrete samples. We applied the Jenson-Shannon divergence (JSD) to evaluate quantitatively the temporal change in the PDFs because it is a finite value, compared with other information divergence such as Kullback-Leibler divergence (KLD). The JSD is given by:

$$JSD(p_t \mid p_{t-1}) = \frac{1}{2} KLD(p_t \mid q) + \frac{1}{2} KLD(p_{t-1} \mid q) \quad (1)$$

where $p$ indicates the PDF of vital signs and its subscript indicates time; $q$ and KLD are defined as:

Table 1: Patient information.

| | Patient 1 | Patient 2 |
|---|---|---|
| Sex | Female | Male |
| Age (years) | 94 | 100 |
| Disease | Cerebral infarction | Pneumonia |

$$q = \frac{1}{2}(p_t + p_{t-1}) \quad (2)$$

$$KLD(p_t \mid p_{t-1}) = \sum_i p_t(i) \log \frac{p_t(i)}{p_{t-1}(i)}. \quad (3)$$

When the two PDFs, $p_t$ and $p_{t-1}$, are more similar, the JSD decreases. By determining the temporal change in the JSD, we can identify the characteristic changes in the patient's status.

# 3 Results

## 3.1 Temporal Changes in the PDFs of Vital Signs

Figure 1 shows the temporal changes in the PDFs of the heart and respiratory rates of patients 1 (a) and 2 (b), with the horizontal and vertical axes indicating the elapsed time and vital signs, respectively. The color shows the probability or frequency of the data. Figure 2 explains the interpretation of Figure 1, where the three continuous lines indicate the probability functions with different variance (large, medium, and small) and the colored figures next to each probability function indicate each probability function represented by a color map. In the color maps, the reddish and bluish colors represent higher and lower probabilities, respectively. Thus, the changes in the trend (the most probable values) and variance in time-series data can be easily visualized. In patient 1 (Figure 1 (a)), the most probable value of the heart rate fluctuated from 40 to 70 days and then remained constant after 100 days. Regarding the respiratory rate, although the PDF had large variance until 100 days, the variance decreased after 110 days. Figure 3 compares the PDF before 110 days (60 days) with that after 110 days (160 days). Using the PDFs, represented by color maps, the one-dimensional PDF is easily visualized. Although the PDF at 60 days did not have any peaks and had larger variance, the PDF at 160 days showed typical Gaussian distribution with one peak. These color maps enabled the visualization of the temporal change in the PDFs, making it easy to confirm the changes. In patient 2, although both vital signs fluctuated slightly until 90 days, the most probable values drastically increased just 7 days before death. This drastic increase in vital signs can be used as an indicator for impending death.

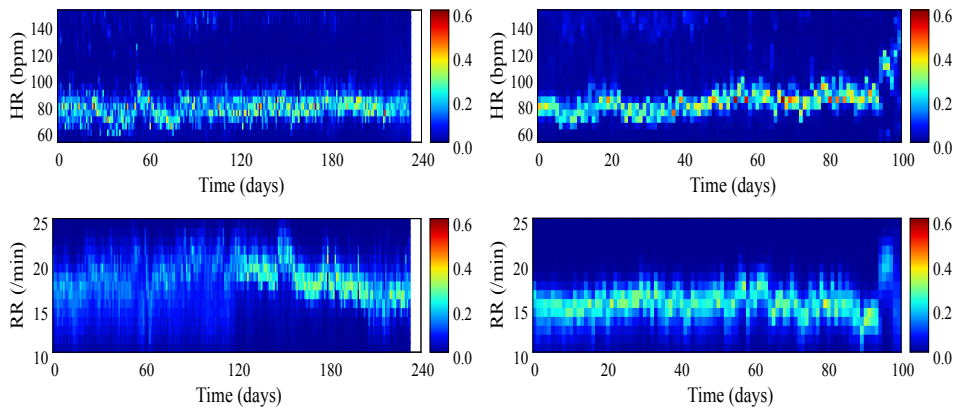
## 3.2 Temporal Changes in the JSD and Missing Data Rate

The temporal change in the JSD of patients 1 and 2 are shown in Figures 4 and 5, respectively. The JSD value is a measure normalized between 0 and 1, and when two PDFs are closer, the JSD gets smaller. Although no specific changes in the temporal JSD of patient 1 were identified, a drastic change was noted in patient 2 a few days before death. The JSD of vital signs could be useful information, at least in patient 2, in estimating the last days of the life.

The missing data rate in time-series data is shown in Figures 4 and 5. The missing data rate of patient 1 exponentially increases a few weeks before death and can be an indicator for impending death. For patient 2, however, the missing data rate does not show characteristic changes before death like in patient 1. Therefore, the missing data rate cannot be a universal indicator for impending death. Table 2 summarizes the applicability of the JSD and missing data rate as signs for predicting death in the two patients.

# 4 Discussion

We have studied the detection of characteristic changes in vital signs and their association with impending death based




(a) Patient 1 (b) Patient 2

Figure 1: Temporal changes in the PDFs of vital signs.

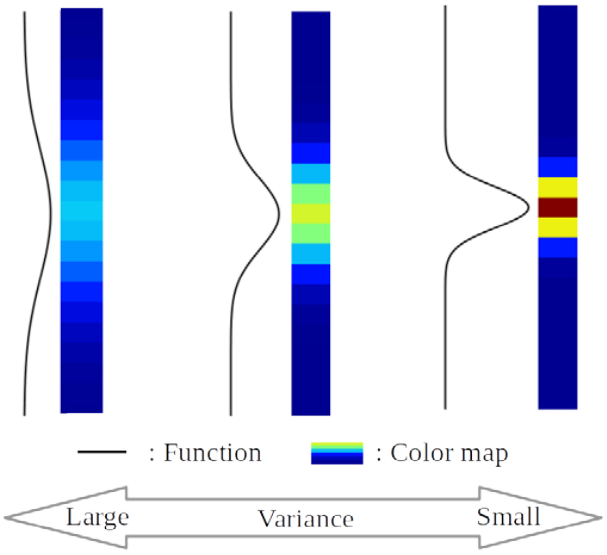



Figure 2: How to interpret the PDFs represented by color maps.

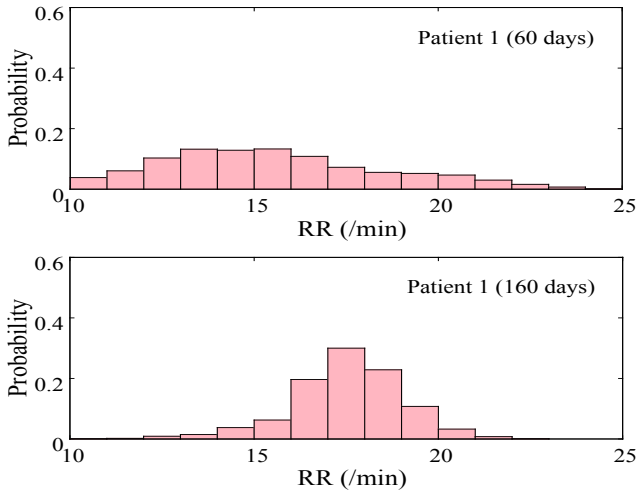



Figure 3: PDFs of the respiratory rate of patient 2 at 60 and 160 days.

on the data measured by the pressure-sensing mat. We found a significant increase in the missing data rate in the final 2 weeks of life for patient 1 and a drastic increase in the JSD in the last 3 days of life for patient 2. However, we could not identify the universal characteristic changes in vital signs related to impending death.

A number of researches studied how vital signs vary at the end of life. To our knowledge, this study is the first attempt to utilize high-temporal-resolution data measured by a less invasive pressure-sensing mat and to analyze large data using a machine learning method to examine the variations in the final days of the patient's life. Although Bruera et al. [4] do not support routine vital sign monitoring of patients who are imminently dying because vital signs cannot be used for prognosis prediction, we believe that it would be acceptable with less invasive sensors such as the pressure-sensing mat used in this study.

However, there are certain limitations in our study. First, the data utilized for the analysis were from only two subjects. Second, we were not able to define the criterion or threshold value to detect the changes associated with impending death because of limitation of data. More data should be analyzed to demonstrate the practicability of the proposed method. Future study needs to investigate the applicability of other sophisticated method for time-series analysis such as seasonal adjustment method [17].

## 4 Conclusion

We have proposed a new methodology to detect specific changes in vital signs that are related to impending death based on the high-temporal-resolution data measured by a pressure-sensing mat and data analysis using density ratio

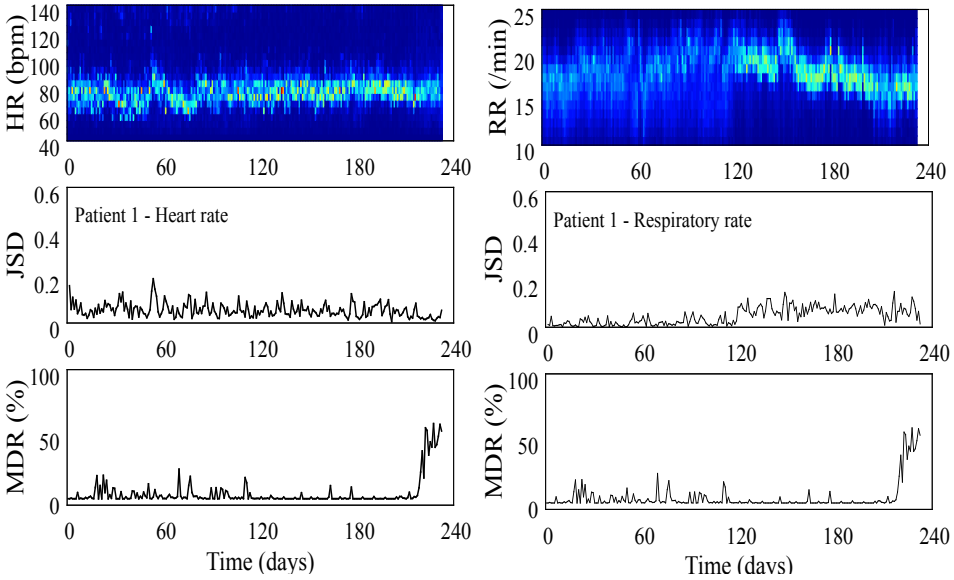



Figure 4: JSD and missing data rate of patient 1.

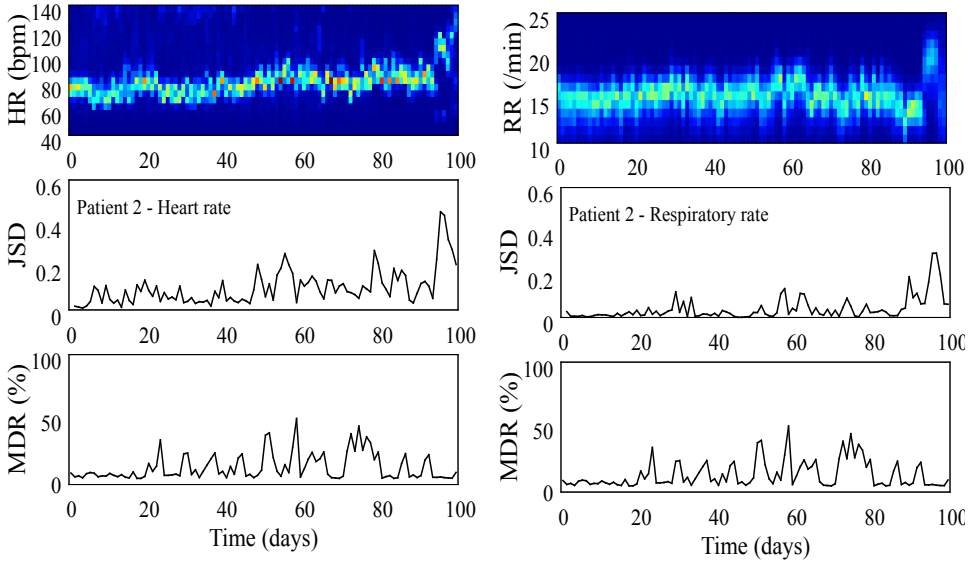



Figure 5: JSD and missing data rate of patient 2.

Table 2: Applicability of JSD and data missing rate as signs for predicting death for two patients.

| | **Patient 1** | **Patient 2** |
|---|---|---|
| JSD | Not applicable | Applicable |
| Data missing rate | Applicable | Not applicable |

estimation. We have identified the characteristic changes such as a drastic increase in JSD and missing data rate of vital signs. However, they are not universal characteristic changes in the data and have limitations in predicting impeding death with confidence. Further research to collect more data and analysis are needed to identify universal characteristic changes for recognizing impending death.

## 6 Author Contribution

T. Shuku, N. Sakano and M. Morita have contributed to the conception and design of research; N. Sakano and S. Kasahara collected and managed the research data; T. Shuku and M. Mizuki analyzed the research data; T. Shuku, N. Sakano, M. Morita and S. Kasahara performed the interpretation the data and results. T. Shuku has drafted the article; N. Sakano revised the article. All authors have given approval of the final version to be published.

## 7 Conflict of Interest

The authors declare no conflicts of interest associated with this manuscript.

## References


[1] Morita T, Ichiki T, Tsunoda J, Inoue S, Chihara S. A prospective study on the dying process in terminally ill cancer patients. Am J Hosp Palliat Care. 1998; 15: 217-222.

[2] Lunney JR, Lynn J, Foley DJ, Lipson S, Guralnik JM. Patterns of functional decline at the end of life. JAMA. 2003; 289: 2387-2392.

[3] Kao YH, Chen CN, Chiang JK, Chen SS, Huang WW. Predicting factors in the last week of survival in elderly patients with terminal cancer: a prospective study in southern Taiwan. J Formos Med Assoc. 2009; 108: 231-239.

[4] Bruera S, Chisholm G, Santos RD, Crovador C, Bruera E, Hui D. Variations in vital signs in the last days of life in patients with advanced cancer. J. Pain Symptom Manage. 2014; 48: 510-517.

[5] Hui D, Santos RD, Chisholm G, Bansal S, Crovador CS, Bruera E. Bedside clinical signs associated with impending death in patients with advanced cancer: preliminary findings of a prospective, longitudinal cohort study. Cancer. 2015; 15: 960-967.

[6] Saliba D, Elliott M, Rubenstein LZ, Solomon DH, Young RT, Kamberg CJ. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. J Am Geriatr Soc. 2001; 49: 1691-1699.

[7] Porock D, Parker-Oliver D, Petroski GF, Rantz M. The MDS Mortality Risk Index: the evolution of a method for predicting 6-month mortality in nursing home residents. BMC Res Notes. 2010; 3: 200.

[8] Yourman LC, Lee SJ, Schonberg MA, Widera EW, Smith AK. Prognostic indices for older adults: a systematic review. JAMA. 2012; 307: 182-192.

[9] Nishida Y, Takeda M, Mori T, Mizoguchi H, Sato T. Unrestrained and non-invasive monitoring of human's respiration and posture in sleep using pressure sensors. JRSJ. 1998; 16: 705-711.

[10] Watanabe K, Watanabe H. Study on the non-restrictive vital bio-measurement by the air mattress method. SICE. 2000; 36: 894-900.

[11] Kogure T, Shirakawa S, Shimokawa M, Hosokawa Y. Automatic sleep/wake scoring from body motion in bed: validation of a newly developed sensor placed under a mattress. J Physiol Anthropol. 2011; 30: 103-109.

[12] Pal A, Mukherjee A, Dey S. Future of healthcare – sensor data-driven prognosis. In: Prasad R, Dixit S, editors. Wireless world in 2050 and beyond: a window into the future! Springer Series in Wireless Technology. Switzerland: Springer; 2016. p. 93-109.

[13] Sugiyama M, Suzuki T, Kanamori T. Density ratio estimation in machine learning. Cambridge: Cambridge University Press, 2012.

[14] Rakthanmanon T, Campana B, Mueen A, Batista G, Westover B, Zhu Q. Searching and mining trillions of time series subsequences under dynamic time warping. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 12; 2012 Aug 12-16; Beijing, China: ACM; 2012.

[15] Idé T, Inoue K. Knowledge discovery from heterogeneous dynamic systems using change-point correlations. In: Proceedings of 2005 SIAM International Conference on Data Mining (SDM 05), Newport Beach, CA, 571-576, 2005.

[16] Gombay E. Change detection in autoregressive time series. J Multivariate Anal. 2008; 99: 451-464.

[17] Genshiro K. Introduction to time-series modeling. Boca Raton, Florida: CRC Press, 2010.

# A Hybrid Electronic Health Record System Integrating Electronic and Paper-based Records

**Kenshi Terajima[1,4]*, Norito Negishi[2], Kouichi Maruyama[3], Hiroaki Hasegawa[2] and Kouhei Akazawa[1]**

[1] Department of Medical Informatics, Niigata University Medical and Dental Hospital, Niigata, Japan

[2] FINDEX Inc., Japan

[3] Management Planning Division, Niigata University Medical and Dental Hospital, Niigata, Japan

[4] Department of Medical Informatics, Uonuma Institute of Community Medicine, Niigata University Medical and Dental Hospital, Niigata, Japan

## Abstract

**Objectives:** Ordinary electronic health record (EHR) systems have difficulty managing paper-based medical records because digitized paper-based records must be browsed with a viewer other than the main display for the EHR system. Because of this "fragmentation" of electronic and paper information, it is difficult to combine keyboard/mouse input with handwriting input in a clinical setting. We developed a new EHR system to address this problem.

**Methods:** We revised a commercial EHR system to function as a progress note system that integrates keyboard/mouse-based electronic records with digitized paper-based records, both of which can be viewed in a page-turning style. In this system all records, including paper-based records, are arranged in relation to the time they were written, even when they were digitized later at a scanning center. These features of the system allow for combined use of handwritten and keyboard/mouse input, without fragmentation. We investigated the impact of the introduction of the first EHR system in the ophthalmology department at our hospital, which had long resisted implementation of an EHR system.

**Results:** The number of ophthalmology outpatients did not significantly decrease after introduction of the system. The ophthalmologists in our hospital accepted the first EHR system without negative reactions, and a gradual transition from handwritten input to keyboard/mouse input was noted.

**Conclusions:** The present system enables users to produce flexible medical documentation, by using both keyboard/mouse and handwriting inputs, without reducing medical efficiency or safety. This is a cost-efficient, true hybrid digital–analog EHR system.

### Keywords

Electronic health records; Paper-based record; Handwriting; Ophthalmology

**Correspondence to:**

**Kenshi Terajima, M.D., Ph.D.**
Department of Medical Informatics, Uonuma Institute of Community Medicine, Niigata . University Medical and Dental Hospital, Niigata, Japan.
E-mail: terajima@bri.niigata-u.ac.jp

**EJBI 2018; 14(1):58-66**
Received: December 18, 2017
Accepted: February 14, 2018
Published: February 21, 2018

## 1 Introduction

Adoption of electronic health record (EHR) systems has been rapidly accelerating because of the potential for such systems to improve delivery, quality, and efficiency of health care [1]. However, a fundamental shortcoming of EHR systems is the difficulty in managing paper-based medical records such as handwritten medical charts, interview sheets, referral forms by other medical institutions, and informed consent forms with patient signatures [2]. Paper-based records are usually digitized by a scanner and browsed with a viewer other than the main display for the EHR system [3]. In other words, keyboard/mouse-based electronic records and paper-based records cannot be viewed in the same way, a situation referred to as "fragmentation" of electronic and paper information (Figure 1). Because this situation reduces the efficiency and safety of health care, it is difficult to combine keyboard/mouse input with handwriting input in clinical settings [4].

Some doctors still prefer classical handwriting input to keyboard/mouse input in medical documentation and

are reluctant to use an EHR system. As a method of producing narrative progress notes, handwritten input has definite advantages over keyboard/mouse input. Keyboard input is slower than handwriting [5, 6]. In addition, keyboard/mouse input records are inherently more restrictive as compared with handwritten sketches or diagrams on paper; therefore, keyboard/mouse input progress notes often have fewer graphic representations [5, 6, 7, 8]. Electronic records are more likely to include copied and repetitive notes [5, 9]. Some reports indicate that constant use of a keyboard and a mouse results in less eye-to-eye contact between doctors and patients [4, 10, 11]. Therefore, some physicians consent to the adoption of an EHR system only if it permits handwriting on paper.

C-Note (FINDEX Inc., Tokyo, Japan) is a progress note system developed to respond to such doctors' requests, in which various types of medical records can be viewed in a page-turning style [12]. On the pages we can type text, draw illustrations, import data and images from external testing equipment, and make annotations on an image by using a mouse or a pen-tablet device as if we are writing on paper. Although this system offers various methods for inputting medical documentation with a paper-like interface, fragmentation of electronic and paper information remains a problem (Figure 2). A simple solution to this problem is immediate scanning of paper-based records, followed by manual pasting of the digitized images onto a progress note page during the patient's care. This is an unrealistic solution in the busy setting of outpatient care because timely manual manipulation of records by doctors and clinical clerks is not always possible. Additionally, in large hospitals, paper-based records are often collected and digitized at a scanning center at least several hours, or possibly a day, after the records are created. In such cases, keyboard/mouse-based electronic records and digitized paper-based records are sometimes not arranged in the order they were written, which can result in miscommunication of patient information among medical staff.

## 2 Objectives

To address the problems described above, we developed a new EHR system by revising C-Note. The new system integrates keyboard/mouse-based electronic records with digitized paper-based records, which can both be viewed in page-turning style. All records, including paper-based records, are arranged in relation to the time they were written, even when they were later digitized at a scanning center. Finally, the system allows for combined use of handwritten and keyboard/mouse input, without fragmentation of electronic and paper information. To verify the effectiveness of the system, we investigated the impact of introducing the first EHR system to the ophthalmology department at our hospital.

## 3 Methods

### 3.1 Revision of the C-Note System

In our hospital, MegaOakHR (NEC Corp., Tokyo, Japan)

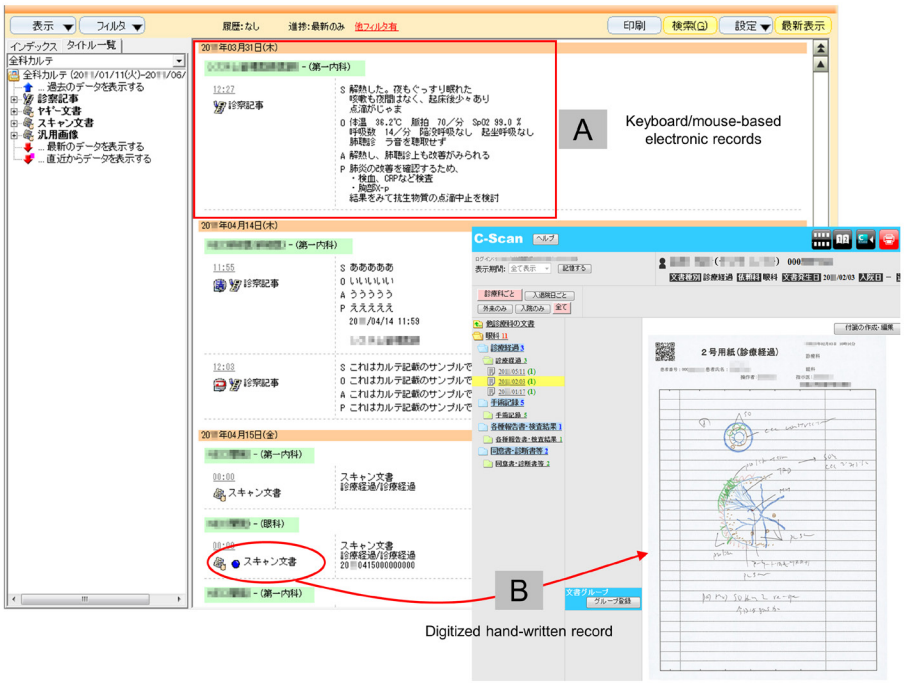



Figure 1: Example of "fragmentation" of electronic and paper information.
When using an ordinary EHR system, keyboard/mouse-based electronic records are browsed in the main system display (A). However, a separate display is needed in order to view digitized paper-based records (B).

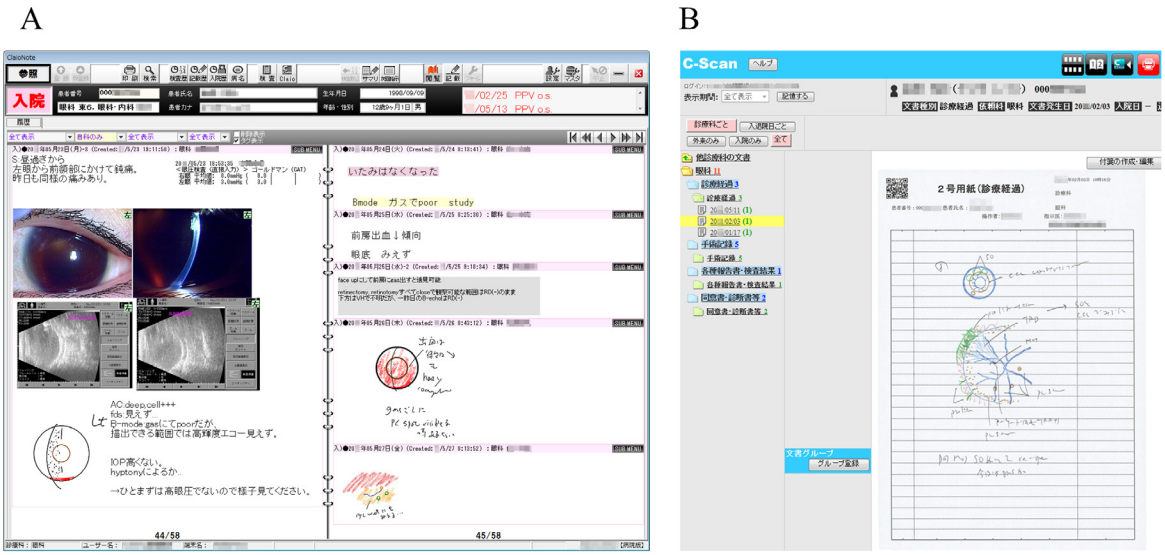



Figure 2: Original progress note system.
C-Note (FINDEX Inc.) is a progress note system in which medical records can be created and then browsed in a page-turning style (A). However, the original version requires a separate display to view digitized paper-based records (B).

is used as the main EHR system, including a computerized provider order-entry system in all departments [13]. Just before writing a paper-based record, the user must print a page with a QR code for the patient, which is generated by the main system. The documents written by the user are later picked up, taken to a scanning center, and then digitized on the same day or the next morning. The QR code contains patient information, the name and department of the user, the time of printing (approximately equal to the time of writing), and document type (eg, progress note, results of laboratory tests, certificates). Use of QR codes ensures that data from digitized paper-based records are registered to the database of all scanned documents in our hospital.

We modified the C-Note system so that all data from a digitized paper-based record are also automatically registered in the system database. Each paper-based record corresponds to one page of the system display. Every time the system database is updated (including modifications or deletions of already registered data), the registered data are rearranged in the order of writing, and the double-page view of the system is reconstructed. Therefore, keyboard/mouse-based electronic records and paper-based records are arranged in the order they were written and can be browsed chronologically in a page-turning style (Figure 3). To ensure that records remain arranged in the order they were originally created, already registered paper-based records cannot be edited, e.g., by adding annotations.

## 3.2 Adoption of the First Ophthalmic EHR in Our Hospital

Adoption of an EHR system in an ophthalmology department is one of the most demanding challenges in medical record management, because of the unique characteristics of ophthalmic outpatient care, which include the need for many intradepartmental examinations, the necessity of documentation that emphasizes graphical representation of examination findings, unique outpatient workflows, and high clinical volumes [7, 14, 15, 16]. An ophthalmic medical record may contain various data types, such as texts, drawings, photographs, images, graphs, waves, and schematic diagrams. Additionally, the time for writing such records is limited because, as compared with other departments, more patients are evaluated and the durations of evaluations are shorter. Therefore, the ophthalmology department in our hospital had long declined EHR implementation, although several EHR systems specially designed for ophthalmic medical record keeping had been proposed [17, 18]. However, we repeatedly explained the unprecedented features of the revised C-Note system to the department, which finally agreed to introduce the system—the first ophthalmic EHR system in our hospital.

To examine how the system was used, we calculated the ratio of the numbers of paper-based to keyboard/mouse-based records. All medical records produced by 28 ophthalmologists during the 8 months after the system was introduced were categorized as keyboard/mouse-based or paper-based records. Then, the numbers of paper-based and keyboard/mouse-based records were counted. The count ratios of keyboard/mouse-based to paper-based records (keyboard/paper) were monitored during these 8 months. We

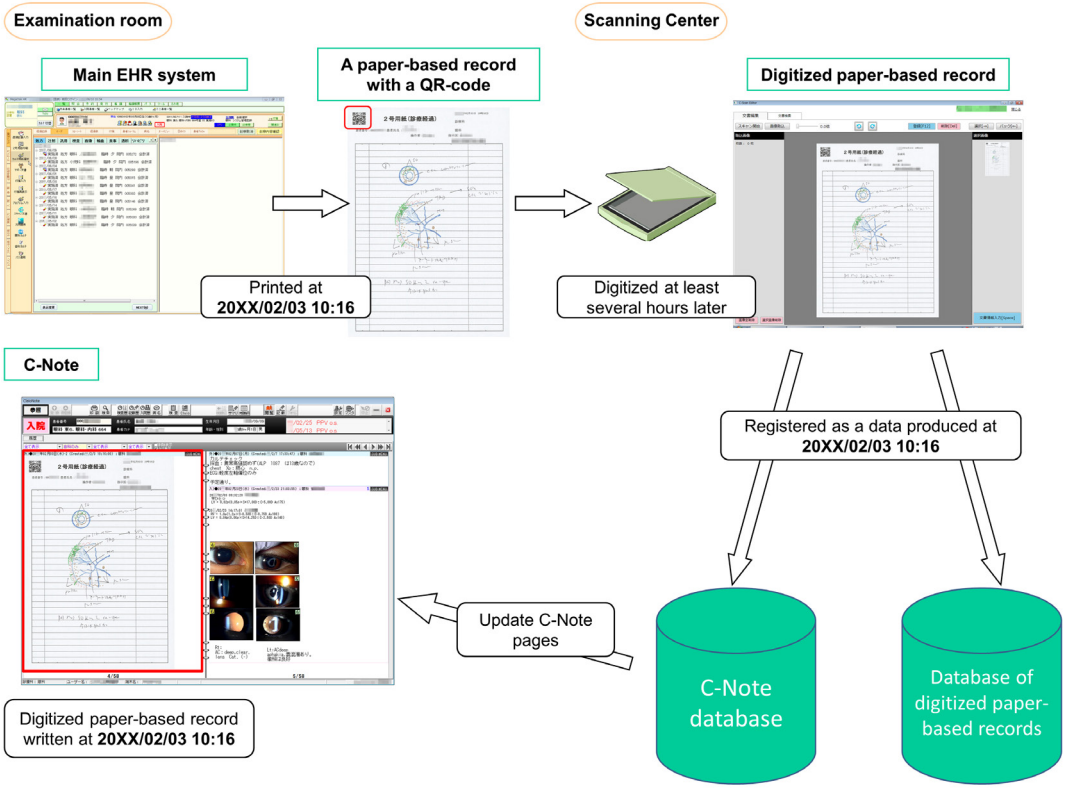



Figure 3: The revised progress note system integrates electronic and paper-based medical records.

When a paper-based record with a QR code is later digitized at a scanning center, the digitized data are also automatically registered to the system database as data produced at the time of writing. The double-page view is reconstructed every time the system database is updated. Therefore, keyboard/mouse-based electronic records and digitized paper-based records are arranged in the system in the order they were written.

also counted the numbers of ophthalmic outpatients seen during the periods 6 months before and 6 months after implementing the system.

## 4 Results

Before implementing the system, the department used only handwritten paper-based medical records, including more than 20 types of ophthalmic charts, interview records, and informed consent forms. When the system was introduced, we converted these paper-based documents to templates with QR codes stored in the main EHR system. Representative templates are shown in Figure 4. In the system, both keyboard/mouse-based electronic records and digitized paper-based records can be browsed in the same page-turning style (Figure 5).

Therefore, a doctor sometimes used these paper-based templates for handwritten input and other times produced records with a keyboard/mouse input. Figures 6 and 7 shows the actual proportions of the usage of these two input methods during the 8 months after system implementation. As shown in Figure 6, the count ratio of keyboard/mouse-based to paper-based records increased from 1.58 to 2.02 during the 8-month period, which suggests a gradual transition from handwriting input to keyboard/mouse input. However, the count ratio varied widely by doctor (Figure 7). Some doctors mostly used keyboard/mouse input from the beginning (Drs. A and B in Figure 7; a transient decrease in the keyboard/paper ratio for the two doctors in June was due to a temporary increase in their request for digitizing documents of new outpatients referred by other medical institutions), while other doctors predominantly used handwritten input (Drs. F and G in Figure 7) throughout the observation period. Despite differences in uptake among physicians, our results show that the ophthalmologists accepted the new EHR system without negative reactions, which resulted in a gradual transition from handwritten input to keyboard/mouse input.

Figure 8 shows the number of ophthalmic outpatients seen per day during the 6 months before and after introduction of the system. Although there was an almost 10% decrease in the number of outpatients during the 1–2 months after introducing the system, the number returned to the usual level at 3 months after the system was introduced. The average numbers of ophthalmic outpatients per day 6 months before and after implementing the revised system were 122 and 114,

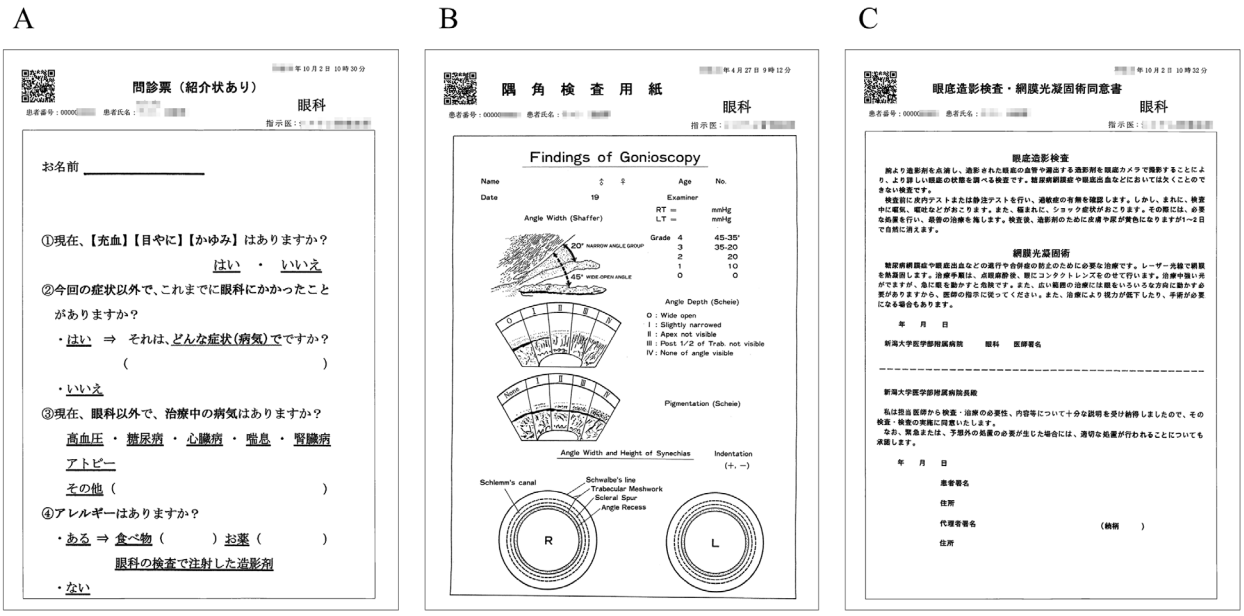



Figure 4: Representative templates (with QR codes) used in our hospital for paper-based records.

When the revised progress note system was introduced, we registered to the system more than 20 types of handwritten paper-based templates with QR codes for ophthalmic charts, interview records, and informed consent forms. The three representative templates shown are an interview record for a new patient (A), a record of an ophthalmic examination (B), and an informed consent document (C).

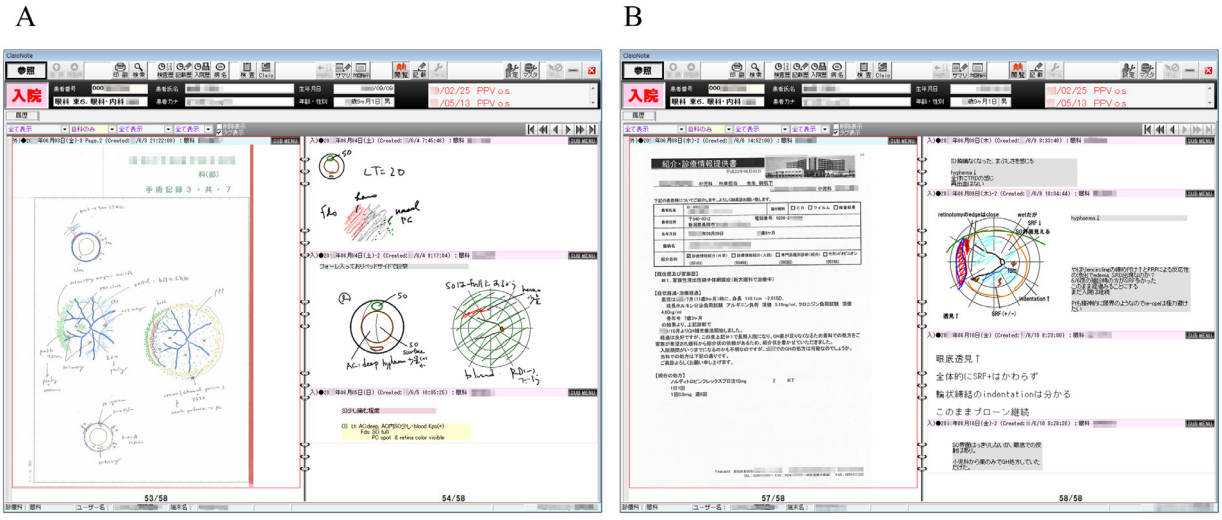



Figure 5: Representative double-page views from the revised progress note system.

In this system, both keyboard/mouse-based electronic records and digitized paper-based records can be browsed in a page-turning style. The examples from the revised progress note system are a handwritten record of a surgical procedure and an electronic record (A) and a paper-based referral form from another hospital and an electronic record (B).

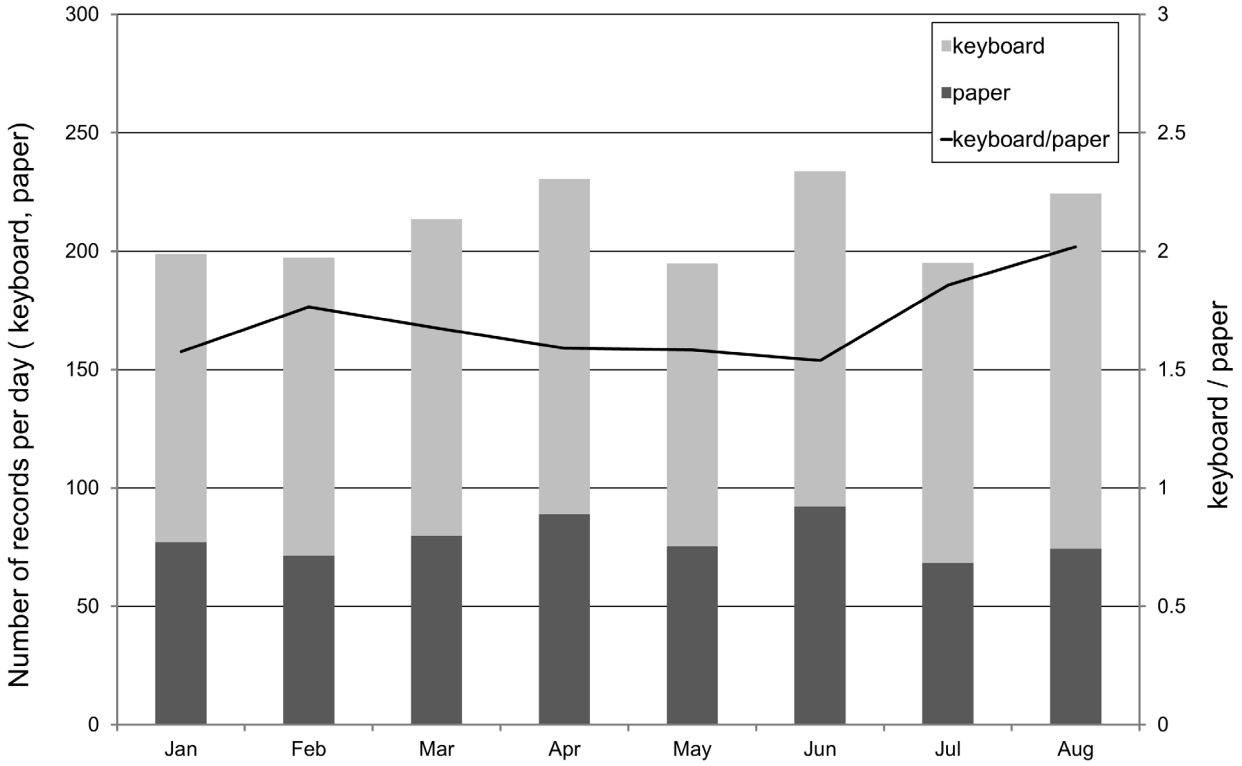



Figure 6: The numbers of keyboard/mouse-based (keyboard) and paper-based (paper) records and their count ratio (keyboard/paper). The keyboard/paper ratio was calculated by using data collected from 28 ophthalmologists during the 8-month period after the system was implemented. The ratio increased from 1.58 to 2.02 during the 8-month observation period, which suggests a gradual transition from handwritten to keyboard/mouse input.

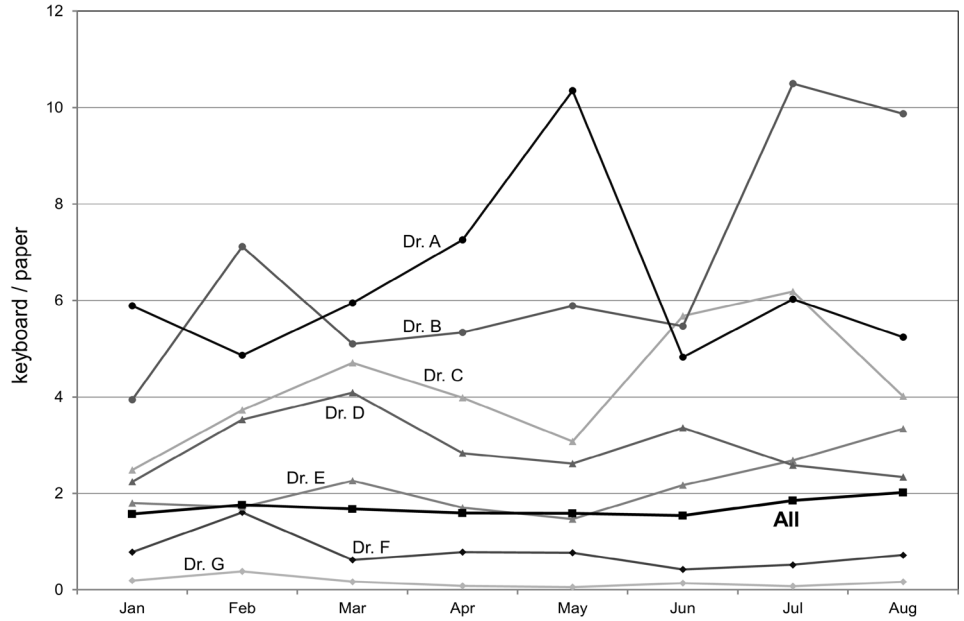



Figure 7: Count ratios of keyboard/mouse-based to paper-based medical records (keyboard/paper) for eight representative doctors.

Keyboard/paper ratio varied widely among the physicians. Some doctors used mostly keyboard/mouse input from the beginning (Drs. A and B), while others predominantly used handwritten input (Drs. F and G) throughout the observation period.

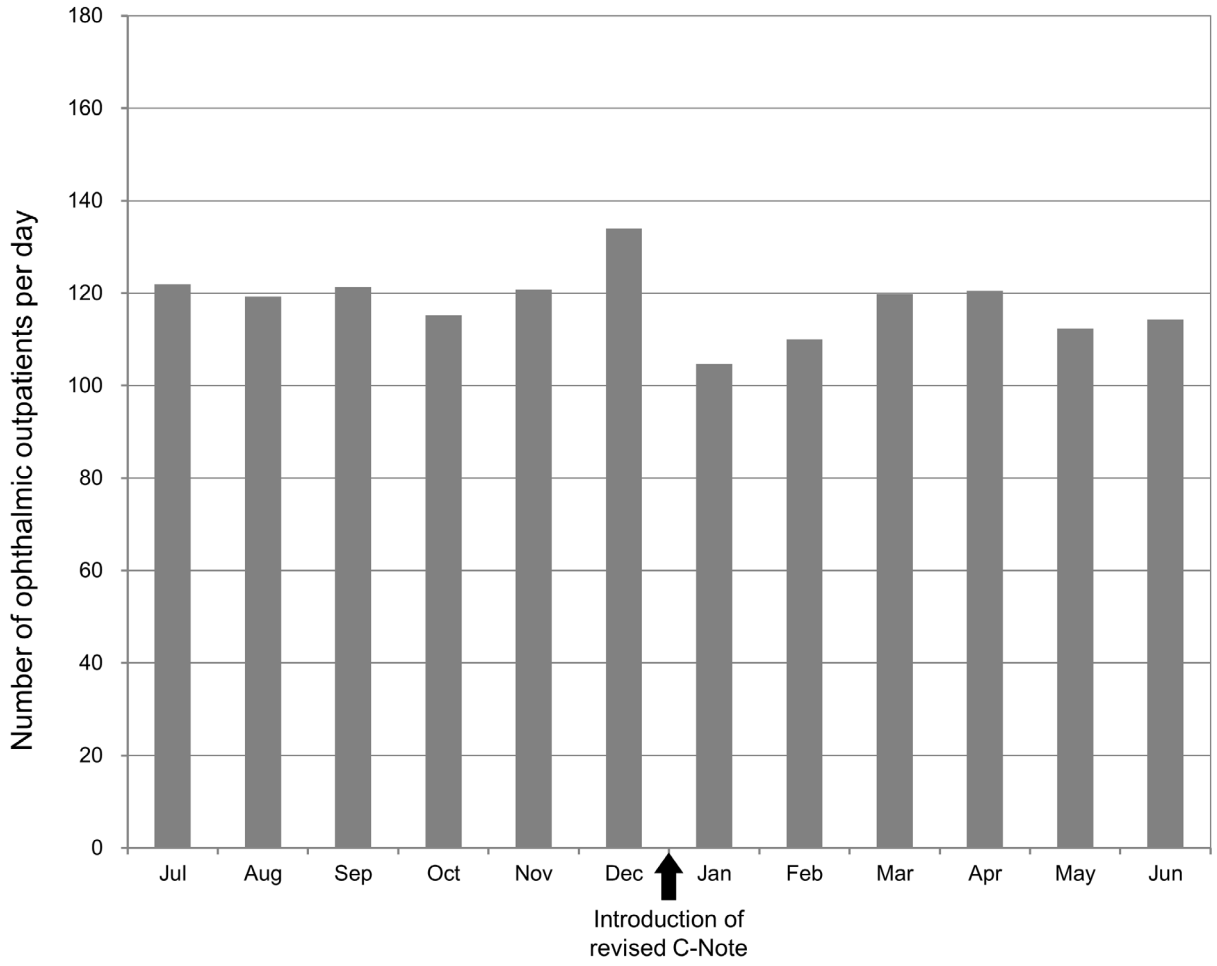



Figure 8: Number of ophthalmic outpatients per day 6 months before and after introduction of our system.

Although there was an almost 10% decrease in the number of outpatients during the 1–2 months after introducing the system, the number had increased to normal volume at 3 months after system introduction.

respectively. These results indicate that there was a minimal and transient decrease in the number of outpatients seen per unit of time after implementing the system.

## 5 Discussion

The present results show that the first ophthalmic EHR implementation in our hospital was successful. The decrease in clinical volume was minimal and transient; thus, the clinical outflow in the ophthalmology department was hardly affected by EHR implementation. That contrasts strikingly with previous studies reporting the opposite effect [19, 20]. The ophthalmologists accepted the new EHR system, and the system utilization rate increased in relation to their satisfaction with the system. Nevertheless, most ophthalmologists still have concerns regarding the potential adverse effects on productivity and efficiency in ophthalmology practice [18, 19, 21]. Furthermore, a recent survey found that ophthalmologists' rate of satisfaction with their EHR has decreased [17]. We believe that the success in our hospital is strong evidence of the effectiveness of our revised EHR system.

The most important advantage of our system is the coexistence of electronic and paper-based records. The system offers complete browsing compatibility for the two types of records, which are arranged in the order they were written rather than by the time of registration. Therefore, in this system, classical handwriting in paper is accepted as an equivalent input method to keyboard/mouse input, and there is no communication gap between electronic and paper information. We believe that this is the best solution for doctors reluctant to use an EHR system. Even users who are unfamiliar with the system, e.g., part-time and temporary employees, can utilize the system by mainly using handwriting input from the first day. The transition from handwritten to keyboard/mouse input can then occur gradually, in accordance with users' computer literacy and understanding of the system. Users can take advantage of the respective benefits of handwritten and keyboard input.

Similar systems or devices that aim for coexistence of electronic and paper-based records have important practical

limitations. First, electronic document management systems (EDMSs) are software programs that manage the creation, storage, and control of documents electronically and can integrate scanned documents [22, 23]. However, unlike an EHR system, an EDMS is only used for reference and is not suitable for creating progress notes. Second, to our knowledge, only one other EHR system (MegaOak-NEOCIS, NEC Corp., Tokyo, Japan [24]) developed for large Japanese hospitals is based on concepts similar to those guiding the development of our system. However, the initial cost of the MegaOak-NEOCIS is at least five times that of our system. Furthermore, our system allows for easy handling of medical records from ophthalmology, otorhinolaryngology, and gynecology departments, without a supplementary system. Third, in contrast to digital pens, digital writing boards, and tablets or sheets of exclusive paper optimized for scanning, our system does not require additional devices or supplies, the ongoing costs of which are usually much higher than expected, especially for a large hospital such as ours. Therefore, we think our system is the most feasible choice in a general hospital.

As discussed above, in the Introduction, similar results can be achieved by clinical clerks who scan paper by hand and insert it into the appropriate place of an EHR. However, this method will likely limit flexibility regarding when and where doctors write records. Furthermore, it requires many additional workers in a large hospital, where multiple doctors use the EHR system simultaneously. A future controlled study that compares the original system with on-site processing and our revised system with posteriori automatic registration would help confirm the benefits of the latter. We hope this limitation will be addressed soon.

## 8 Conclusion

The present system enables users to produce flexible medical records with keyboard/mouse and handwriting input methods, without fragmentation of electronic and paper information. This system is a cost-efficient, true hybrid digital–analog EHR system.

## 9 Acknowledgement

We thank Toru Hasegawa, Yasuhisa Wakatsuki, Takuo Goto, and Kozue Saito for their dedicated support during the introduction of the system.

## 10 Conflict of Interest

K Terajima recieved lecture fees from FINDEX Inc.

## 11 Author Contribution

K. Terajima has contributed to the conception and design of research and drafted the article, N. Negishi and H. Hasegawa have developed the system, K. Maruyama and H. Hasegawa have analyzed the results of the research and K. Akazawa has approved the system introduction and supervised the research. All authors have given approval of the final version to be published.

## References

[1] Yoshida Y, Imai T, Ohe K. The trends in EMR and CPOE adoption in Japan under the national strategy. Int J Med Inform. 2013; 82: 1004-1011.

[2] Fenz S, Heurix J, Neubauer T. Recognition and privacy preservation of paper-based health records. Stud Health Technol Inform. 2012; 180: 751-755.

[3] Laerum H, Karlsen TH, Faxvaag A. Effects of scanning and eliminating paper-based medical records on hospital physicians' clinical work practice. J Am Med Inform Assoc. 2003; 10: 588-595.

[4] Saleem JJ, Russ AL, Justice CF, Hagg H, Ebright PR, Woodbridge PA. Exploring the persistence of paper with the electronic health record. Int J Med Inform.. 2009; 78: 618-628.

[5] Chan P, Thyparampil PJ, Chiang MF. Accuracy and speed of electronic health record versus paper-based ophthalmic documentation strategies. Am J Ophthalmol. 2013; 156: 165-72 e2.

[6] Chiang MF, Read-Brown S, Tu DC, Choi D, Sanders DS, Hwang TS. Evaluation of electronic health record implementation in ophthalmology at an academic medical center (an American Ophthalmological Society thesis). Trans Am Ophthalmol Soc. 2013; 111: 70-92.

[7] Chiang MF, Boland MV, Brewer A, Epley KD, Horton MB, Lim MC. Special requirements for electronic health record systems in ophthalmology. Ophthalmology. 2011; 118: 1681-1687.

[8] Lim MC, Patel RP, Lee VS, Weeks PD, Barber MK, Watnik MR. The long-term financial and clinical impact of an electronic health record on an academic ophthalmology practice. J Ophthalmol. 2015; 329819.

[9] Hartzband P, Groopman J. Off the record--avoiding the pitfalls of going electronic. N Engl J Med. 2008; 358: 1656-1658.

[10] Linder JA, Schnipper JL, Tsurikova R, Melnikas AJ, Volk LA, Middleton B. Barriers to electronic health record use during patient visits. AMIA Annu Symp Proc. 2006; 499-503.

[11] Shachak A, Reis S. The impact of electronic medical records on patient-doctor communication during consultation: a narrative literature review. J Eval Clin Pract. 2009; 15: 641-649.

[12] http://findex.co.jp/en/pdf/cnote_en.pdf

[13] Namikawa H, Miyakawa R, Sato Y, Takashima K. Electronic Medical Record System "MegaOakHR". Nec Tech J. 2008; 3: 89-93.

[14] Chiang MF, Boland MV, Margolis JW, Lum F, Abramoff MD, Hildebrand PL. Adoption and perceptions of electronic health record systems by ophthalmologists: an American Academy of Ophthalmology survey. Ophthalmology. 2008; 115: 1591-1597; quiz 7 e1-5.

[15] Matsuo T, Gochi A, Hirakawa T, Ito T, Kohno Y. Outpatients flow management and ophthalmic electronic medical records system in university hospital using Yahgee Document View. J Med Syst. 2010; 34: 883-889.

[16] Park JSY, Sharma RA, Poulis B, Noble J. Barriers to electronic medical record implementation: a comparison between ophthalmology and other surgical specialties in Canada. Can J Ophthalmol. 2017; 52: 503-507.

[17] Boland MV, Chiang MF, Lim MC, Wedemeyer L, Epley KD, McCannel CA. Adoption of electronic health records and preparations for demonstrating meaningful use: an American Academy of Ophthalmology survey. Ophthalmology. 2013; 120:1702-1710.

[18] Sanders DS, Read-Brown S, Tu DC, Lambert WE, Choi D, Almario BM. Impact of an electronic health record operating room management system in ophthalmology on documentation time, surgical volume, and staffing. JAMA Ophthalmol. 2014; 132: 586-592.

[19] Lam JG, Lee BS, Chen PP. The effect of electronic health records adoption on patient visit volume at an academic ophthalmology department. BMC Health Serv Res. 2016; 16:7.

[20] Redd TK, Read-Brown S, Choi D, Yackel TR, Tu DC, Chiang MF. Electronic health record impact on productivity and efficiency in an academic pediatric ophthalmology practice. J AAPOS. 2014; 18: 584-589.

[21] Read-Brown S, Hribar MR, Reznick LG, Lombardi LH, Parikh M, Chamberlain WD. Time Requirements for Electronic Health Record Use in an Academic Ophthalmology Center. JAMA ophthalmol. 2017; 135: 1250-1257.

[22] Kohn D. Electronic document management systems: an overview. Topics in health information management. 2002; 23: 1-6.

[23] Schmidt RA, Simmons K, Grimm EE, Middlebrooks M, Changchien R. Integration of scanned document management with the anatomic pathology laboratory information system: analysis of benefits Am J Clin Pathol. 2006; 126: 678-683.

[24] http://www.nec.co.jp/press/ja/0601/1702.html

# Searching for "Neuromarkers" Characteristic for Pathologic Changes in Schizophrenia by Using the Scaling Indices of the Cerebral Bioelectric Activity

**Oleg Yu. Mayorov[1,2,4*] and Vladimir N. Fenchenko[1,2,3]**

[1] Kharkiv Medical Academy of Postgraduate Education of the Ministry of Health, Ukraine

[2] Institute for Medical Informatics and Telemedicine, Kharkiv, Ukraine

[3] B.Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences, Kharkiv, Ukraine

[4] Research Institute for Children and Adolescents Health Protection of the National Academy of Medical Sciences, Kharkiv, Ukraine

## Abstract

**Introduction**: The study is devoted to the development of new analytical procedures that would be fit for further study of the cerebral bioelectric activity (EEG) and may become reliable tools for diagnosing schizophrenia by using EEG.

**Objective of the study:** The study was aimed at detecting new "neuromarkers" suitable for diagnosing schizophrenia by using EEG.

**Materials and methods:** The subjects of the study were healthy individuals and patients with schizophrenia, both groups having been examined while either in the state of the undisturbed wakefulness or under mental exertion (backward counting in memory). Fractal characteristics of EEG traces has been determined by applying the Multifractal Detrended Fluctuation Analysis (MDFA).

**Results:** The dimensionality values of the dominant monofractal (the localization of the multifractal spectral function maxima) in patients with schizophrenia while in the state of undisturbed wakefulness could be increased in comparison with the healthy subjects. Have introduced a "decision-making rule" that allows to rate the status of tested subjects with special reference to the risk factors (suspected signs of schizophrenia) by employing the sets of data obtained by applying the procedures of multifractal analysis towards the EEG traces.

A diagnostic method allowing the detection of EEG signals pathologically deviating from normal ranges of values based on the use of multifractal characteristics has been suggested.

## Keywords

EEG; Multifractality; Multifractal detrended fluctuation analysis; Multifractal spectrum; Classification; Nearest *k*; Neighbor algorithm method

**Correspondence to:**

**Oleg Yu. Mayorov**
Institute for Medical Informatics and Telemedicine, Kharkiv, Ukraine.
E-mail: o.y.mayorov@gmail.com

**EJBI 2018; 14(1):67-74**
Received: November 20, 2017
Accepted: February 14, 2018
Published: February 21, 2018

## 1 Introduction

Despite that numerous [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] attempts have been made to analyze the bioelectric activity in patients with schizophrenia, no reliable criteria for evaluating EEGs were developed [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. EEG traces display a variety of diverse abnormalities, the degree of which is not expressed clearly and which are not significantly correlated with the degree of psycho-pathologic syndromes [10, 11]. Doubtless, this disorder is associated with functional disintegration of cortical mechanisms ("*a disconnection hypothesis*") [6], and it could be the underlying cause for other mental afflictions similar to schizophrenia [6, 7, 8, 9, 10, 11]. In other words, not merely separate systems or structures are getting impaired, but their cooperative integrated functionality itself, which results in the onset of schizophrenia. All of the above calls for the development of new analytical procedures that would be fit for further study of the cerebral bioelectric activity and may become reliable tools for diagnosing schizophrenia by using EEG traces.

The processes that occur within the brain, find their reflection in EEG signals, while their structure is complex and multifractal. Such processes can be quantified by way of factorization into domains with variable local scale characteristics and the peculiar spectra of scaling parameters. Processes of this kind have been named multifractal - as opposed to monofractal ones, whose scaling parameters remain invariant respective to the scale of the given domain.

Basic concepts of the multifractality theory have been formulated by Mandelbrot B [12, 13] and were shaped into a rigorous theory by Parisi G and Frisch U [14], Benzi R, et al. [15].

A wide-spread application of the multifractal analysis method towards arrays of biomedical data began following the publications by P.S. Ivanov et al. [16, 17], who have provided an experimental evidence for the fact that multifractal spectra of the human cardiorhythm could be efficiently used for diagnosing a variety of disorders.

Time-related dynamics of scaling patterns can be studied by a variety of approaches. Classical analytical procedures are applicable *only towards parameters of stationary processes*, whereas biological processes are distinguished by the heterogeneity and are not stationary.

Among the most efficient tools fit for the task of processing time series produced by such processes are the Wavelet Transform Modulus Maxima – WTMM [18, 19] and the method of the Multifractal Detrended Fluctuation Analysis - MDFA [20, 21, 22, 23].

Many contemporary studies have so far been focused primarily on the comparative statistical evaluation of the fractal parameters quantified by using these two methods [20, 24].

The majority of authors [18, 19, 20, 21, 22, 23], however, were employing simulative models based on the use of extended signals. In such cases, the randomly sampled measured values approximated the computed data sufficiently well. Meanwhile, time series of data obtained in medical-biological experiments are much less extended. The tool of preference for such series is the MDFA, which is, in fact, a kind of a variance analysis procedure intended for analyzing one-dimensional stochastic fluctuations [22]. The latter procedure allows for assessing the effects of prolonged correlations in non-stationary time series and provides arrays of fractal parametric values with more statistical regularity than those obtained by applying the WTMM. The correct selection of the signal domain within which the pattern of the scale variation would allow for the wavelet transformation, on the other hand, is a difficult task. If the assessed interval is relatively short, then, the values obtained by applying the WTMM procedure could reflect false multifractal properties [24, 25].

## 2 Materials and Methods

The study involved the group of healthy subjects (the reference group - 35 male patients aged 21-23) and the group of patients with schizophrenia that received no treatment (the trial group - 28 male patients aged 21-26, whose diagnosis has been confirmed clinically). The subjects in both groups have been examined while in the states of undisturbed wakefulness or under mental exertion (backward counting in memory).

EEG traces were obtained by applying the monopolar acquisition technique adjusted to the 24-channel electroencephalograph "DX-systems" (Ukraine), (Certificate No 008/17/MD, Validity: to 20.06.2022, Ukrainian Certification Agency® (UCA®) Accredited by NAAU 10310 IS0/IEC 17065)) equipped with the averaged reference electrode ("10-20 system" according to Goldman and Offner). This version of the EEG acquisition technique provides an opportunity for tracking the cerebral activity locally - in the zones beneath the electrodes.

## 3 Results

### 3.1 Formulation of the Problem. Analytical Methodology

A study of the fractal characteristics pertaining to the time series of data representing the bioelectrical activity of the human brain (EEG traces) was conducted by applying the MDFA procedure. Diagnosing patients with schizophrenia, particularly at the initial stages of the disorder is quite a difficult task.

According to our initial assumption, the regulatory neurophysiological mechanisms would, under the normal circumstances, produce the dynamic signals that were going to be multifractal by nature, i.e. the multifractality would be further regarded as the "intrinsic" property of normally functioning control entities. Hence, measuring either the degree of the depletion of the multifractality or the variability of its parameters could provide invaluable diagnostic information concerning the status of the brain and its responses towards external stimuli.

EEG signals registered by one or other lead are quite complex structurally and are the results of the superimposition of a variety of signals originating from diverse sources (cerebral structures), which could be performing specific functions not necessarily correlated at a particular given point in time.

In clinical practice, only several specific frequency domains within the EEG signal patterns happen to deserve the attention of an examining specialist ($\delta$, $\theta$, $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2$, $\gamma$). The activity within each of the domains reflects the status of the brain and is associated with certain cerebral systems and mechanisms.

The factorization of the EEG signal into the component rhythms by applying the frequency filters would not, however, take into account a very peculiar phenomenon, namely, the

presence in EEG signals of incorporated fragments (subsets) that are scale-invariant.

Scale invariance presupposes the invariance of fundamental characteristics of specific fragments in a situation of the scale transition; that is to say, the structure of a fragment on one given time scale would remain similar to its structure on an alternative scale. Notably, the time scale shift is possible within certain limits only, while the structure of the resulting pattern is not identical to that of the original fragment; statistical characteristics of both, however, do coincide.

Thus, the fine structure of EEG signals appears to contain characteristic spectra of fractal dimensions; based on these, an attribute vector intended for identifying cerebral conditions (including the status at the initial stages of schizophrenia) could be put in place.

Temporal sequences of data obtained by registering bioelectric potentials of the cerebral cortex, which, in fact, are arrays of values of the EEG signals measured at singular leads at equal time intervals, have been conventionally named "incremental" series.

The use of an MDFA procedure version adapted for such series is very well documented; its application towards the EEG signals, nevertheless, calls for specific modifications due to the fact that such signals contain rhythms of variable frequencies, amplitudes and, what is essential, of variable functional significance. The latter circumstance should always be taken into account whenever any kind of meaningful interpretation of multifractality of measured EEG signals is being attempted.

One fundamental step that the application of the MDFA towards the incremental series does involve is the extraction of a fluctuation profile from the initial series by using computed mean values. Further, this profile is expanded into non-overlapping segments each with the length of *s* (as the length of the series is not always divisible by *s*, the number of points contained in the last domain is less than *s*, therefore, it is recommended that the expansion procedure be repeated, but, this time, from the opposite end of the sequence). Next, a local polynomial trend is determined for each of the segments and variance $\sigma_n(s)$ is assessed, where $n = 1, 2, ..., N_s$ is the number of the segment. Then, a so-called "deformed" variance.

$$\sigma(q,s) = \left\{ \frac{1}{N_s} \sum_{n=1}^{N_s} \sigma_n(s)^q \right\}^{\frac{1}{q}}$$

calculated by raising a variance value for each segment to a power $q$ with a subsequent group averaging for all segments (at $q = 0$ the following equation should be used):

$$\sigma(0,s) = e^{\frac{1}{2N_s} \sum_{n=1}^{N_s} \ln \sigma_n(s)}$$

If the analyzed series could be reduced to a self-similar set with long-range correlations, then the "deformed" variance $\sigma(q,s)$ within the wide range of values $s$ could be presented as an exponential function $\sigma(q,s) \approx s^{h(q)}$

where $h(q)$ is the generalized Hurst exponent, or scaling exponent.

In time series correspondent to a monofractal set, variance $\sigma_n(s)$ is similar for all segments, while the Hurst exponent is segment-invariant, and the generalized Hurst exponent is not dependent on the deformation parameter $q$.

In multifractal series at positive $q$ the major contributing components of $\sigma_n(s)$ are segments with greater variance levels $\sigma_n(s)$, while at negative $q$ the dominant role belongs to segments with lesser variances $\sigma_n(s)$.

In other words, at negative values of $h(q)$ the Hurst exponent $h(q)$ reflects the properties of segments with lesser fluctuations, while at positive values - of those ones with greater fluctuations.

Further, the standard interpretation of the scaling characteristics of a series would require a transition from the Hurst exponent $h(q)$ to the mass exponent [15] $\tau(q) = qh(q) - 1$, which, at $q << 0$ and $q >> 0$ increases linearly in direct proportion to $q$, while in the proximity to $q = 0$, it acquires non-linear properties which causes the impediment of growth of $\tau(q)$ in proportion to the growth of the parameter $q$. In monofractals this domain is altogether absent, while $\tau(q)$ grows linearly and proportionally to $q$.

Mass exponent allows the calculation of the spectral function values $f(\alpha) = \alpha q(\alpha) - \tau(q(\alpha))$, $\alpha = \tau'(q)$, the latter being the primary characteristic of multifractals [15].

The application of the MDFA (incidentally, as much as of the WTMM) towards the empirical values involves interpretational difficulties - in the event if a singularity spectrum $f(\alpha)$ is discrete, as the target of evaluation is the envelope of the local maxima, which would include a set of false points. Nonetheless, the analysis of multifractal EEG signals, which are generated by complex entities, typically produces continuous multifractal spectra that could be approximated by asymmetric spectral function plots, whereas the respective time series are significantly correlated temporally.

The spectral function $f(\alpha)$ defines a set of monofractals with dimensions $\alpha$, which configure a given multifractal. The position of its maximum would determine the dimensionality of a particular monofractal, whose contribution to the statistical summation would be dominant. The latter parameter is one of the most important characteristics of the multifractality of a given time series.

Another significant variable is the width of the multifractal spectra, i.e. the length of the domain within which the function is positive. The values of this function would

increase in a direct proportion to the degree of expression of the multifractality of the series in question.

In all probability, whether in the state of undisturbed wakefulness or while under mental exertion, patients with schizophrenia could display such cerebral activity patterns that may not be normally present in healthy individuals. Therefore, it is not unlikely that dimensions of dominant monofractals in subjects with schizophrenia could be, on the average, greater than in healthy study participants. Predictably, the multifractal spectral widths in schizophrenia should be abnormally greater, the causative factor being the existence of accessory ("coupling fault") connections among the systems that generate EEG signals.

In healthy individuals, the greater complexity of cortical reactions could be anticipated in response to the challenge of a mental exertion; as a consequence, the distribution of spectral function maxima could become right-skewed - hence, the dimensionality of the dominant monofractal may increase.

It could be speculated that in schizophrenic patients, in whose brain the coherent connections between cerebral structures had already been altered, such dislocation of the spectral function maxima could be of a different fashion; as much modified could be the shift at the level of multifractal spectral widths.

Thus, it appears that there is a certain degree of correlation between the dimensional variability of dominant multifractals and the increase of multifractality properties of the series. Thus, the assumption is that the most influential variable might be the value of the dimensionality of the dominant monofractal as measured at the level of a particular lead.

It turns out, then, that dimension values pertaining to dominant monofractals either in the state of undisturbed wakefulness or under the mental exertion could be arranged into series of elements ("neuromarkers"), which may further be used for rating the status of patients who have been a priori hypothetically diagnosed with initial stages of the schizophrenic disorder.

### 3.2 Evaluation of the Parameters Characterizing the Multifractality of the EEG Signal in Healthy and Affected Subjects

The objective of the study was the comparative evaluation of parameters that are employed as measures of multifractality of EEG signals obtained from healthy and affected subjects alike while in the state of undisturbed wakefulness or under the mental exertion (backward counting in memory).

Figure 1 shows a plot representing the characteristic dependency of the *generalized* Hurst exponent upon the deformation parameter (obviously, we get a *regular* Hurst exponent) obtained for a healthy individual in the state of undisturbed wakefulness and in the state of the mental exertion by using frontal leads (F3, F4), while the Figure 2 demonstrates the spectral function curve obtained by using the same leads.

The mental exertion was shown to have resulted in a significant transformation of the generalized Hurst exponent values along with the respective variation of spectral function values (based on use of frontal leads).

Of note: the generalized Hurst exponent values were dependent upon the variations of the deformation parameter both quantitatively and qualitatively. Such dependence revealed itself not only in dislocated spectral function maxima, but also in increased spectral widths. In other words, the greater was the dimensionality of the dominant monofractal the higher was the multifractality level, which, apparently, was suggestive of the added complexity of reactions occurring in the respective areas of hemispheres. At the same time, the frontal leads placed around the right hemisphere (F4) did not register any significant increase of spectral widths, while the degree of the individual variance in all groups was unusually high. This phenomenon is surely explicable, as the left hemisphere is known to contain a so-called "logic perception area" involved in logic-based activities (reading, linguistic perception, mathematical operations (calculus)), while the structures of the right hemisphere are dedicated to the spacial perception, are responsible for the perception of the "general" information, and are possibly capable of functioning in a variety of alternative ways for various individuals.

Figures 3 and 4 show the dependence of the generalized Hurst exponent values upon the deformation parameter variations along with the spectral function values obtained from the a patient with schizophrenia while in the states of undisturbed wakefulness and mental exertion (data obtained by using frontal leads (F3, F4)).

In this case, the values of Hurst exponent were at the significantly higher level than in a healthy subject. This fact suggests that the EEG traces may contain persistent series, which means that signals are positively autocorrelated, i.e. that the buildup of a signal would, most likely, produce a cumulative effect in the future and vice versa.

All of the above implies that the brain of a patient who is affected by the disorder and who is in the state of rest is running the processes different from cerebral reactions of healthy subjects.

The mental exertion, on the other hand, affects the dimensionality of a dominant monofractal so much less significantly, that any tangible advancement in complexity of the relevant cortical functionality could hardly be imagined.

Of note: the variance of values pertaining to the functions

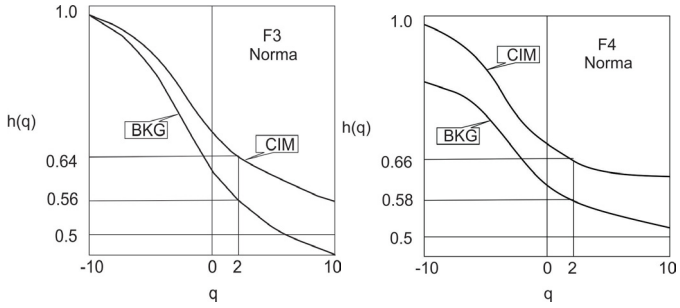



Figure 1: The comparison of generalized Hurst exponent values of a healthy subject in the states of undisturbed wakefulness (BKG) and mental exertion (CIM) (data obtained by using frontal leads (F3, F4)).

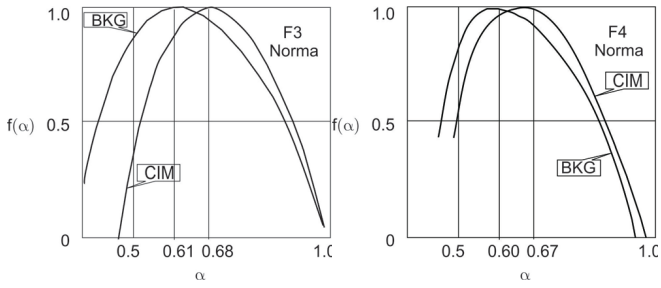



Figure 2: The comparison of spectral function values of a healthy subject in the states of undisturbed wakefulness (BKG) and mental exertion (CIM) (data obtained by using frontal leads (F3, F4)).

in question in patients with schizophrenia was significantly greater than in healthy individuals.

Tables 1 and 2 give calculated mean values of multifractality parameters of EEG signals measured in subjects while in the states of undisturbed wakefulness and mental exertion (the only data represented are sets measured by using the frontal (F3, F4) and parietal (P3, P4) leads registering the integrative activity and by using temporal leads (T3, T4) that provided correlates of emotional reactions).

In healthy subjects the transition towards the mental exertion has caused the dislocation of spectral function maxima, which was right-skewed at the level of all leads, i.e. the mental exertion has somewhat increased the dimensionality of the dominant monofractal.

In patients with schizophrenia the dislocation of the spectral function maximum under the mental exertion was inexplicit or, in some cases, even non-extant. In other words, the cortex of patients affected by the disorder was obviously devoid of the specific type of the connectivity between its structures.

In patients with schizophrenia, while in the state of undisturbed wakefulness, the *dimensionality of the dominant monofractal* (as a measure of the position of the spectral function maximum evaluated at all leads) appeared to have been greater than in healthy subjects (the level of significance being 0.95).

It should be emphasized, that in healthy individuals, even

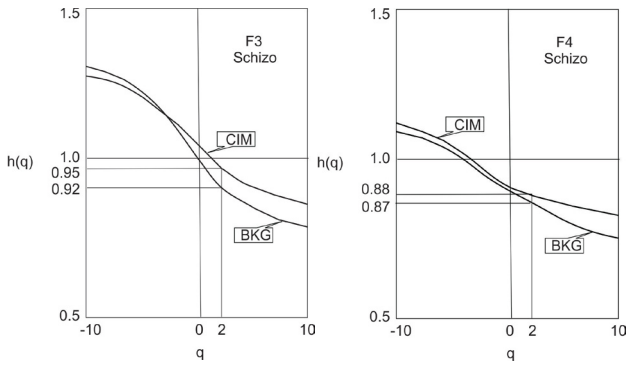



Figure 3: The generalized Hurst exponent values in a patient with schizophrenia in the states of undisturbed wakefulness (BKG) and mental exertion (CIM) (data obtained by using frontal leads (F3, F4)).

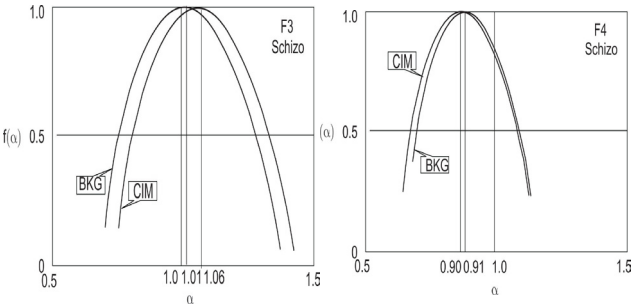



Figure 4: Spectral function values in patients with schizophrenia in the states of undisturbed wakefulness (BKG) and mental exertion (CIM) (data obtained by using frontal leads (F3, F4)).

while under mental exertion, the dimensionality of the dominant monofractal was at the lower level, than in patients with schizophrenia while they were at rest.

The *multifractal spectral* width values in healthy subjects, while they were switching to the mental exertion, have been somewhat growing, which was the evidence for the strengthening of connectivity between the signal-generating systems. In patients with schizophrenia subjected to the mental exertion, by contrast, the multifractal spectrum width values showed no tendency towards increase - much in accordance with our preliminary assumption. It should be noted that, while at rest, the group of patients with schizophrenia did show somewhat greater values of the multispectral width in comparison with the group of healthy subjects, which we explained by the presence of accessory ("coupling fault") connections between the signal-generating entities.

In conclusion, a neurophysiological interpretation of the results may be attempted. From the standpoint of a neurophysiologist, various local scale-sensitive properties of the EEG signals, the spectrum of their scale-related parameters, i.e. a so-called *multifractal spectrum*, reflect a series of rapidly interchanging mental processes. A multifractal EEG signal is an entity inclusive of several *monofractal* processes, which possess scaling characteristics remaining scale-invariant

Table 1: Location of spectral function maximum.

| Leads | Mean value in healthy subjects | | Mean value in patients with the disorder | |
|---|---|---|---|---|
| | at rest (BKG) | under exertion (CIM) | at rest (BKG) | under exertion (CIM) |
| F3 | 0.62 ± 0.02 | 0.69 ± 0.02 | 0.95 ± 0.04 | 0.97 ± 0.09 |
| F4 | 0.60 ± 0.02 | 0.67 ± 0.04 | 0.77 ± 0.05 | 0.81 ± 0.08 |
| T3 | 0.73 ± 0.02 | 0.77 ± 0.02 | 0.85 ± 0.04 | 0.82 ± 0.06 |
| T4 | 0.77 ± 0.02 | 0.82 ± 0.04 | 0.83 ± 0.03 | 0.82 ± 0.07 |
| P3 | 0.56 ± 0.02 | 0.59 ± 0.01 | 0.75 ± 0.05 | 0.76 ± 0.07 |
| P4 | 0.54 ± 0.02 | 0.58 ± 0.01 | 0.69 ± 0.05 | 0.69 ± 0.06 |
| Mean value for all leads combined | 0.64 ± 0.04 | 0.69 ± 0.04 | 0.79 ± 0.05 | 0.79 ± 0.05 |

Table 2: Values of the multifractal spectral width.

| Leads | Mean value in healthy subjects | | Mean value in patients with the disorder | |
|---|---|---|---|---|
| | at rest (BKG) | under exertion (CIM) | at rest (BKG) | under exertion (CIM) |
| F3 | 0.5 0± 0.03 | 0.50 ± 0.05 | 0.68 ± 0.03 | 0.70 ± 0.05 |
| F4 | 0.47 ± 0.03 | 0.53 ± 0.05 | 0.54 ± 0.04 | 0.53 ± 0.03 |
| T3 | 0.51 ± 0.03 | 0.53 ± 0.03 | 0.64 ± 0.04 | 0.58 ± 0.05 |
| T4 | 0.46 ± 0.02 | 0.53 ± 0.04 | 0.55 ± 0.03 | 0.56 ± 0.03 |
| P3 | 0.51 ± 0.03 | 0.51 ± 0.03 | 0.56 ± 0.05 | 0.52 ± 0.03 |
| P4 | 0.51 ± 0.04 | 0.53 ± 0.03 | 0.63 ± 0.05 | 0.51 ± 0.03 |
| Mean value for all leads combined | 0.49 ± 0.10 | 0.52 ± 0.09 | 0.60 ± 0.05 | 0.55 ± 0.07 |

within any given range of scales. Amongst those there are isolated dominant monofractals that are conditioned by prevailing mental reactions.

While in the state of undisturbed wakefulness, a significant difference between healthy test subjects and patients with schizophrenia has been detected - the dimensionality of scaling spectra pertaining to dominant monofractals in healthy individuals was lower than in patients with schizophrenia. This fact supports the assumption that the affected patients, while even in the state of undisturbed wakefulness, do experience effects of numerous mental reactions making them incapable of being relaxed even under favorable circumstances.

While under the mental exertion, the healthy subjects have displayed increased dimensionality of dominant monofractals, whereas the patients with schizophrenia showed no changes whatsoever; the greater dimensionality was detected either at rest or under mental exertion (backward counting in memory). These phenomena are apparently either the reflection of the stochastic nature of cerebral reactions (whether at rest or in the state of mental exertion, regardless) or the results of the impaired integrative functions. The lack of ability to relax or to get focused on the concrete task constitute the causative basis for the situation of the emotional distress.

### 3.3 Metric Classification of Healthy and Schizophrenic Patients by *k*-Nearest Neighbor Algorithm Method

Thereby, we arrive to the confirmation of the hypothesis, according to which the status of the cerebrum of an imaginary subject could be quite precisely described by using a two-dimensional spatial set of characters, where a reference point is selected; the coordinates that define this point are the ***mean values*** of the dimensionality of the dominant monofractal for all leads combined either in the state of undisturbed awakening or under mental exertion. Figure 5 presents the set of such points obtained by studying select healthy subjects and patients with schizophrenia; the points have been plotted by using the following coordinates: "dimensionality values of the dominant monofractal measured at rest" against the "dimensionality values of the dominant monofractal under the mental exertion". It can be clearly seen that the points form the clusters in the vicinity of the mean values, thereby forming compact domains within the sample space. Notably, the clusters of points pertaining to the healthy subjects are quite dense; separation of any point from a cluster, therefore, could be interpreted as a deviation from the norm (by no means, though, should this separation be treated as a diagnostic sign of schizophrenia). Points pertaining to the

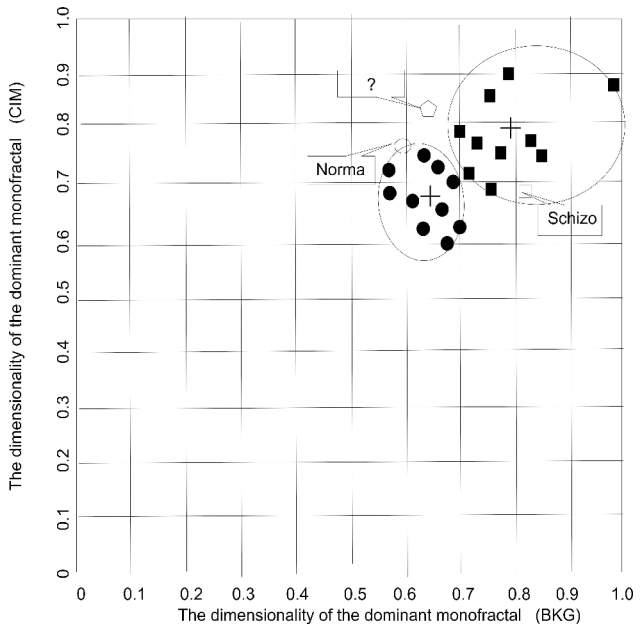



Figure 5: Filled characters correspond to patients belonging to reference groups (bullets - the state of norm, squares - schizophrenia), unfilled characters correspond to specific states of various groups of patients (bullets - the state of norm, squares - the state of schizophrenia is suspected, rhombuses - deviations from the state of norm are present).

patients affected by the disorder are somewhat "scattered" within the system of coordinates, and this variance is fairly natural. Migration of a point into this domain would be suggestive of the state of schizophrenia.

As the contours of domains correspondent to groups of healthy subjects and patients with schizophrenia alike are geometrically unsophisticated, the task of detecting suspected schizophrenia cases can be reduced to the application of metric classification procedures, e.g. of a relatively efficient and simple one $k$ - the nearest neighbor algorithm procedure ($k$-nearest neighbor algorithm, KNN) [26]. This procedure is based upon the comparison of levels of probability with which a given tested individual may be theoretically associated with either a group of subjects with abnormalities unrelated to schizophrenia or with a group of those rated as suspected schizophrenia cases with special focus on the degree of affiliation between either of the groups and the majority of the test subject's nearest neighbors belonging to the "training set".

The distance to every "neighbor" is essential - the smaller the distance, the greater is the effect of the neighbor. Importantly, this algorithm is resilient against abnormal fluctuations and, if the necessity arises, can be easily adjusted in order to ensure the best possible precision of rating [27]. Schematically, the functionality of this algorithm is shown in Figure 5. Filled characters correspond to the patients from reference groups (bullets - healthy subjects, squares - patients with schizophrenia). Unfilled characters correspond to specific states in various groups of subjects (bullets - the state is of norm, square - the state of schizophrenia is suspected, rhombuses - the patient's state is deviating from the norm, though schizophrenia is not suspected)

## 4 Conclusion

The results of the present studies have suggested that the dimensionality values of the dominant monofractal (the localization of the spectral function maxima) in patients with schizophrenia while in the state of undisturbed wakefulness could be increased in comparison with the healthy subjects. On the other hand, the transition towards the state of mental exertion (backward counting in memory) in healthy subjects has resulted in a more or less significant right-skewed dislocation of the spectral function maxima, while in patients with schizophrenia any significant redistribution was not observed.

Based on these observations, the authors have introduced a "decision-making rule" that allows to rate the status of tested subjects with special reference to the risk factors (suspected signs of schizophrenia) by employing the sets of data obtained by applying the procedures of multifractal analysis towards the EEG traces in the states of undisturbed wakefulness and mental exertion.

## 5 Author Contribution

Prof. Oleg Mayorov is the scientific supervisor of the project, selection and EEG examination of healthy and schizophrenics. Computer EEG analysis and interpretation of results. Mathematical support was given by Vladimir Fenchenko.

## 6 Conflict of Interest

The authors declare no conflicts of interest associated with this manuscript.

## References

[1] Gandal MJ, Edgar JC, Klook K, Siegel SJ. Gamma synchrony: towards a translational biomarker for the treatment resistant symptoms of schizophrenia. Neuropharmacology. 2012; 62: 1504-1518.

[2] Castelnovo A, Ferrarelli F, D'Agostino A. Schizophrenia: from neurophysiological abnormalities to clinical symptoms. Front Psychol. 2015; 6: 478.

[3] Ferrarelli F. Sleep in Patients with Schizophrenia. Curr Sleep Medicine Rep. 2015; 1: 150–156.

[4] Ferrarelli F, Riedner B, Peterson MJ, Tononi G. Altered prefrontal activity and connectivity predict different cognitive deficits in schizophrenia. Human Brain Mapping. 2015; 36: 4539–4552.

[5] Kaskie RE, Graziano B, Ferrarelli F. Schizophrenia and sleep disorders: links, risks, and management challenges. Nat Sci Sleep. 2017; 9: 227-239.

[6] Friston KJ. Theoretical neurobiology and schizophrenia. Brain Med Bull. 1996; 52: 644-655.

[7] Woodruff P, Murray R. The aetiology of brain abnormalities in schizophrenia. In: Ancil RJ, Holliday S, Higenbottam J, editors. Schizophrenia: Exploring the Spectrum of Psychosis. Chichester, UK: Wiley; 1994, pp. 95–144.

[8] Andreasen NC. A Unitary Model of Schizophrenia: Bleuler's "Fragmented Phrene" as Schizencephaly. Arch Gen Psychol. 1999; 56: 781-787.

[9] Peled A. Multiple constraint organization in the brain: A theory for schizophrenia. Brain Res Bull. 1999; 49: 245-250.

[10] Coutin-Churchman P, Anez У, Uzcategui M, Alvarez L, Vergara F, Mendez L. Quantitative spectral analysis of EEG in psychiatry revisited: drawing signs out of numbers in a clinical setting. Clin Neurophysiol. 2003; 114: 2294-2306.

[11] Gruzelier JH. Theory, methods and new directions in the psychophysiology of the schizophrenic process and schizotypy. Int J Psychophysiol. 2003; 48: 221-245.

[12] Mandelbrot B. Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence. Statistical Models and Turbulence. 1972; 12: 333–351.

[13] Mandelbrot BB. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. J Fluid Mech. 1974; 62: 331–358.

[14] Parisi G, Frisch U. On the singularity structure of fully developed turbulence. Proceedings of the International School of Physics "Enrico Fermi". 1985; 84–87.

[15] Benzi R, Paladin G, Parisi G, Vulpiani A. On the multifractal nature of fully developed turbulence and chaotic systems. J Physics A: Mathematical and General. 1984; 17: 3521–3531.

[16] Ivanov PC, Nunes Amaral LA, Goldberger AL, Havlin S, Rosenblum MG, Struzik ZR et al. Multifractality in human heartbeat dynamics. Nature. 1999; 399: 461–465.

[17] Ivanov PC, Nunes Amaral LA, Goldberger AL, Havlin S, Rosenblum MG, Stanley HE et al. From 1/f noise to multifractal cascades in heartbeat dynamics. Chaos. 2001; 11: 641–652.

[18] Muzy JF, Bacry E, Arneodo A. Multifractal formalism for fractal signals: the structure function approach versus the wavelet transform modulus maxima method. Phys Rev E. 1993; 47: 875.

[19] Muzy JF, Bacry E, Arneodo A. The multifractal formalism revisited with wavelets. Int J Bifurcat Chaos. 1994; 4: 245.

[20] Kantelhardt JW, Zschiegner SA, Bunde A, Havlin S, Koscielny-Bunde E, Stanley HE. Multifractal detrended fluctuation analysis of non-stationary time series. Physica A. 2002; 316: 87-114.

[21] Kantelhardt JW, Koscielny-Bunde E, Rego HHA, Havlin S, Bunde A. Detecting long-range correlations with detrended fluctuation analysis. Physica A. 2001; 295: 441-454.

[22] Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Choas. 1995; 5: 82-87.

[23] Stanley HE, Amaral LAN, Goldberger AL, Havlin S, Ivanov PCh, Peng CK. Statistical physics and physiology: Monofractal and multifractal approaches. Physica A. 1999; 270: 309–324.

[24] Oswiecimka P, Kwapin J, Drozdz S. Wavelet versus detrended fluctuation analysis of multifractal structures. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics. 2006; 74: 161-203.

[25] Veneziano D, Moglen GE, Bras RL. Multifractal analysis: pitfalls of standard procedures and alternatives. Phys Rev E. 1995; 52: 1387-1398.

[26] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. USA: Springer; 2009.

[27] Berry Michael JA, Gordon Linoff. Data mining techniques: for marketing, sales, and customer relationship management. 2nd ed. USA: Wiley; 2004.

# Some Thoughts about Health Informatics in Africa

**Graham Wright**

Department of Information Systems, Rhodes University, Grahamstown, South Africa

**Correspondence to:**

**Professor Graham Wright**, M.Phil, MBA, DN, Cert Ed, FBCS, CITP, MIASHI, SRN, RMN, Department of Information Systems, Rhodes University, Grahamstown, 6140, South Africa.
E-mail: prof.graham.wright@gmail.com

**EJBI 2017; 14(1):75-76**
Received: December 06, 2017
Accepted: December 06, 2017
published: January 08, 2018

I first started working in Africa 12 years ago and my initial impression was that there seemed to be lots of westernised "Capital" or "Large Cities" across the continent and the "rest" was Africa. There are a few real African large cities but the majority of Africa is urban or rural with a very different way of life compared to Europe or North America. Most of the populace has never used a computer and many have never seen one. Indeed in 2014 urban electrification was estimated at 45% of the population and only 19% had access in rural sub-Saharan Africa. Surely Hospitals use Electronic Health Systems? The answer is a firm yes that in the private sector, mainly insurance based, there are billing and payment systems based on ICD10 that record billable care. Indeed there exist systems in state hospitals, including PACS, largely in the developed cities [1]. The two most pervasive systems across the continent are DHIS and OpenClinic GA.

Murray, Betts [2] state that "The District Health Information System (DHIS), originally developed by the University of Oslo, and including the Health Information Systems Program (HISP) is a widely used public health information system, based in free and open source (FOSS) applications for collecting and sharing aggregated data from health facilities and transmitting it to higher levels in the public health system." DHIS is used in over 40 countries including South Africa, which in 2010 introduced a National System based on DHIS that uses registers exclusively as reporting tools, not as clinical tools. Registers are widely used in Public Health Monitoring and Evaluation, which has led to the development of data silos in the majority of countries using them which are of little use to clinicians. In Kenya all health programs are now integrated into DHIS2 for routine reporting.

OpenClinic GA is an open source integrated hospital information management system covering management of administrative, financial, clinical, laboratory, x-ray, pharmacy, meals distribution and other data. Some of its features include a scheduling system, 3BT clinical thesaurus with validated coding aid for ICD-10 and ICPC-2, Snomed CT coding, laboratory order entry and results management, X-ray and pathology, pharmacy stock management and it integrates RxNorm based multi-lingual drug-drug interaction detection. It has installations in the following African countries: Burundi, Cameroon, Congo-Brazzaville, and Democratic Republic of the Congo, Gabon, Ivory Coast, Kenya, Mali, Nigeria, Senegal, Rwanda, Uganda, Sri Lanka, Tanzania and Zambia.

OpenClinic is a complete Hospital Information Management System (HIMS) that is Open Source, well supported and operational in some of the most troubled countries of Africa, some of which are HELINA and IMIA members. Marc Nyssen and Frank Verbeke, from Belgium, also help develop local solutions for Technical Support, HI Associations and Education and Training.

So what has all this got to do with Europe? DHIS2 is developed by HISP which is a global south-south-north collaborative network aiming at better health care in developing countries through the combination of research on and implementation of health information systems coordinated by the Department of Informatics at the University of Oslo (HISP UiO).

OpenClinic GA is developed by staff from BISI: Department of Biostatistics and Medical Informatics, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Belgium and the Belgian Technical Cooperation, Brussels, Belgium and has a base in Burundi.

South Africa has, like other African countries, leap frogged the computer age with mobile phones; an estimated 92 million hand sets owned by a population of 62 million.

So why can we not use mobiles for Health? Access to the Internet is enjoyed by people in first world countries where internet access is cheap and readily available but not so in Africa where cost of data, power outages, education barriers, infrastructure and slow connection speeds are just some of the many challenges facing the mass roll out of internet on the continent. The Internet World Stats for Africa 2016 shows that only 9.3 per cent of people across the African continent are Internet users. Mobile phones ownership is growing in Africa regions in mobile phone use and Internet access, successfully sidestepping the era of desktop computers and landlines.

However many users only access free services for example there is a "call back" which is used to ask someone with airtime to call them, as they have no credit on their "pay as you go" phone. At locations that have free Wi-Fi students connect their mobile phones to use What's App for communication between themselves and Facebook and Instagram for social media [3]. The problem is that data is so expensive in Africa compared to other parts of the world. In addition the minimum wage in South Africa is 3000 rand, which is 178 Euro and the majority of students live on half of that amount. A 500MB data bundle costs 10 Euro or 150 Rand that will buy breakfast for 5 students! Technically Mobile Smart phones could be the answer for those privileged enough to be able to afford Internet access in Africa.

Globally it would, at first glance, appear that mobile phone health apps have a potential for supporting healthcare, especially in Africa, where mobile phones are more prevalent than desktops. However there are challenges that include non-integration with the

health care delivery systems, lack of evidence of clinical effectiveness, and sustained user base. In December 2017 [4] a systematic review was published which looked to find scientific evidence on the efficacy of apps in promoting healthy lifestyles. It came to the conclusion that overall, the evidence so far showed a modest efficacy of apps in health promotion.

Unfortunately I know of no major success stories for the use of mobile phone based health related apps in Africa although there are claims. On examination the apps are either in development or can only be used to connect to specific programs running on a dedicated server.

## References

[1] Wright G, O'Mahony D, Cilliers L. Electronic health information systems for public health care in South Africa: a review of current operational systems. Journal of Health Informatics in Africa. 2017; 4: 51-57.

[2] Murray PJ, Betts HJ, Wright G, Tshayingca-Mashiya NV. Health informatics education and capacity building in Eastern Cape Province, South Africa. Yearb Med Inform. 2009; 158-63.

[3] Wright G, Cilliers L, Niekerk EV, Seekoe E. The Next Stage of Development of Elearning at UFH in South Africa, In: 1th International Conference e-Learning. Lisbon, Portugal, 2017.

[4] Covolo L, Ceretti E, Moneda M, Castaldi S, Gelatti U. Does evidence support the use of mobile phone apps as a driver for promoting healthy lifestyles from a public health perspective? A systematic review of Randomized Control Trials. Patient Educ Couns. 2017; 100: 2231-2243.