A Structural Graph-Coupled Advanced Machine Learning Ensemble Model for Disease Risk Prediction in a Telehealthcare Environment

Raid Lafta, Ji Zhang, Xiaohui Tao, Yan Li, Mohammed Diykh and Jerry Chun-Wei Lin

Abstract The use of intelligent and sophistic technologies in evidence-based clinical decision making support have been playing an important role in improving the quality of patients' life and helping to reduce cost and workload involved in their daily healthcare. In this paper, an effective medical recommendation system that uses a structural graph approach with advanced machine learning ensemble model is proposed for short-term disease risk prediction to provide chronic heart disease patients with appropriate recommendations about the need to take a medical test or not on the coming day based on analysing their medical data. A time series telehealth data recorded from patients is used for experimentations, evaluation and validation. The Tunstall dataset were collected from May to October 2012, from industry collaborator Tunstall. A time series data is segmented into slide windows and then mapped into undirect graph. The size of slide window was empirically determined. The structural properties of graph enter as the features set to the machine learning ensemble classifier to predict the patient's condition one day in advance. A combination of three classifiers—Least Squares-Support Vector Machine, Artificial Neural Network, and Naive Bayes—are used to construct an ensemble framework to classify the graph features. To investigate the predictive ability of the graph with the ensemble

R. Lafta · J. Zhang (✉) · X. Tao · Y. Li · M. Diykh
Faculty of Health, Engineering and Sciences, University of Southern
Queensland, Toowoomba, Australia
e-mail: ji.zhang@usq.edu.au

R. Lafta
e-mail: raidluaibi.lafta@usq.edu.au

X. Tao
e-mail: xtao@usq.edu.au

Y. Li
e-mail: yan.li@usq.edu.au

M. Diykh
e-mail: mohammed.diykh@usq.edu.au

J. C.-W. Lin
School of Computer Science and Technology, Harbin Institute of Technology
Shenzhen Graduate School China, Shenzhen, China
e-mail: jerrylin@ieee.org

classifier, the extracted statistical features were also forwarded to the individual classifiers for comparison. The findings of this study shows that the recommendation system yields a satisfactory recommendation accuracy, offers a effective way for reducing the risk of incorrect recommendations as well as reducing the workload for heart disease patients in conducting body tests every day. A 94% average prediction accuracy is achieved by using the proposed recommendation system. The results conclusively ascertain that the proposed system is a promising tool for analyzing time series medical data and providing accurate and reliable recommendations to patients suffering from chronic heart diseases.

1   Introduction

The chronic diseases such as heart disease have developed and become one of the major public health problems which accounting for 50% of disease burden worldwide [22]. According to World Health Organization (WHO), these diseases were caused more than 60% of all death in 2005 [1]. Nowadays, many of people around the world are suffering from different chronicle diseases because the lack of used diseases prediction tools. Therefore, the survival rates have been noticeably increased due to using sophisticated techniques to predict diseases in a right time.

Recommendation systems are computer-based information systems designed to support and assist medical practitioners in implementation evidence-based practices and improved decision-making [10, 31]. The recommendation systems can help in minimizing medical errors and providing more detailed data analysis in shorter time [38].

Telehealth systems offer a real time and quick way that is enable healthcare practitioners and chronic diseases patients to exchange information easily [11, 45], and subsequently have enjoined fast developments in many countries due to fast service delivery and its low-cost. Most telehealth services are conveyed through Web-based applications which utilize Internet and Web browsers, together with sensors, wearable devices and mobile. Given the significance of disease risk prediction in the medical field [48] as well as the urgency of acquiring more effective analytic techniques for disease risk prediction, great endeavors are expected to enhance the quality of evidence-based decisions and recommendations in the telehealth environment. In telehealth system, patients with chronic heart disease require taking daily medical tests to monitor their heart health conditions. Yet, carrying out various necessary medical tests every day for chronical disease patients in the current practice brings lots of inconvenience and even burden to the patients and adversely affect their life quality. Generating accurate intelligent recommendations to guide their daily medical tests can significantly decrease their workload in taking those tests while keeping the associated health risk in a worthy low level.

In many cases, an accurate medical recommendation is based upon the prediction of patients' short-term disease risk, which is one of the most important functions in telehealth systems. A set of disease risk prediction models have become available in the medical literature using statistical analysis tools and approaches based on data mining tools. These models have been utilized for different healthcare and medical issues [7, 9, 17, 21, 30, 33, 36, 39, 46, 47]. However, most of the existing work only focus on the long-term medical prediction. Nevertheless, the short-term prediction, which is studied in our work, has turns to be more challenging than the long-term prediction as patients' conditions may experience more dramatic and abrupt changes during the short-term timeframe.

In this work, we utilize a structural graph to process the time series medical data of heart disease patients to facilitate the subsequent data analytics to produce the accurate prediction and recommendations. Graphs can be mathematically defined as abstract representations of networks that consist from set of nodes linked by edges [28, 35]. In last years, graph theory has been widely increasingly used in analysing and classification of the complex networks relationships such as, social networks, biological and brain networks, signal and image processing. It is used in neuroscience research to analyse and study the brain diseases [32, 43, 44]. Some studies [16, 26, 37] showed that graph theory can be considered as a one of robust tools to characterize the functional topological properties of brain networks for both normal and abnormal brain functioning [25, 41]. It is also used in image processing as a powerful tool to analyse and classify digital images [34]. The time series of EEG signals are converted into graphs by [12, 13] for EEG sleep stages classification.

The intelligent, accurate medical recommendations in our work rely on the use of classification approaches to produce reliable prediction of the short-term medical risks of the patients. By nature, this is a classification problem which involves using classification methods (called classifiers) to predict the necessity of taking body test of a given medical measurement.

There are several reasons that pushed us to construct the ensemble classifier. First, it provides an efficient solution for building a single model for applications of which the amount of data may be very large [40]. Second, it has also been proven to be an effective tool thanks to its ability to improve the overall accuracy of the prediction model. Empirical results showed that machine learning ensembles are often more accurate than the individual classifiers that make them up [3, 42]. Bagging aggregation is a machine learning ensemble algorithm designed to enhance the accuracy and stability of machine learning algorithms [8], which was proposed by Breiman in the mid-1990s [40]. It has been proven to be a very popular, efficient and effective method for building an ensemble model.

Due to the ensemble outperforms individual classifiers, a combination of three classifiers—Least Squares-Support Vector Machine, Artificial Neural Network, and Naive Bayes—are used to construct an ensemble framework in this work.

The contributions of this work can be summarized as follows:

- First, the time series medical data of a given patient will be segmented into smaller overlapped sliding windows based on the size of the sliding window used in the data analysis;
- Then, each sliding window is mapped into undirect graph in order to extract the structural properties of each graph;
- Finally, the extracted structural properties of each graph are then input into our ensemble learning model to produce a binary recommendation concerning whether that patient needs to take a medical test on the coming day for a certain medical measurement such as the heart rate or blood pressure.

In this paper, a novel short-term recommendation system for chronic heart disease patients is proposed. This system is developed using a structural graph with a machine learning ensemble model to provide patients in a telehealth environment with appropriate recommendations for the necessity of taking a medical body test on the coming day. Such recommendations are established based on the prediction of their heart conditions using their time series medical data from the past few days.

To verify the performance of the proposed model, the metrics of accuracy, workload saving and risk are used and experimental evaluations are conducted on a real-life time series data collected from a pilot study on a group of heart failure patients. The experimental results demonstrate that the proposed model yields a reasonably good recommendation accuracy and can effectively reduce the workload required in medical tests for the patients. It also can effectively reduce the risk of incorrect recommendations. We believe that this analytic model is promising in risk assessment and management associated with heart failure and other similar diseases.

The remainder of this paper is organised as follows. Section 2 explains the details of the proposed methodology including describes the machine learning classifiers that constructing the proposed ensemble model. Section 3 discussed in details the experimental evaluation results and the used dataset and also compared the results of the proposed method with other results of common methods. Finally we conclude the paper and highlight the future work in Sect. 4.

2 Methodology

Figure 1 illustrates the overall architecture of our recommendation system used for chronic heart disease patients in the telehealth environment. In this section, we present in details the architecture of the recommender system.
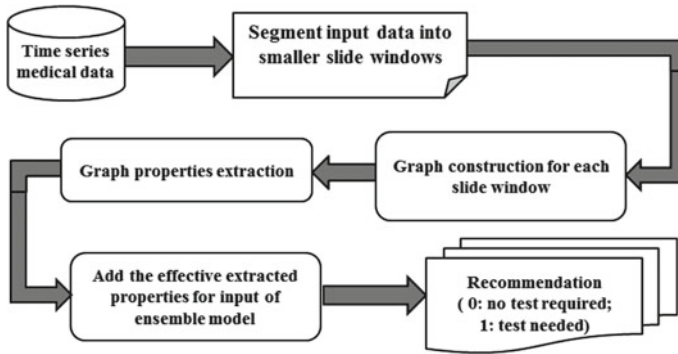
Fig. 1 An overview of the proposed methodology

2.1 Time Series Segmentation

In our system, the input time series data, represented as $X = \{y_1, y_2, y_3, \ldots, y_n\}$ which contains n data, is segmented into a set of overlapped sub-segments based on a predefined value of parameter k that specifies the size of the sliding window, corresponding to each sub-segment. In this work, many experiments are conducted with different numbers of slide window sizes (k). It is important to divide the time series data into several windows because each slide window will be mapped into a separated graph and then extract the effective features from graphs to represent the slide windows.

2.2 Graph Construction and Structural Graph Similarity

Each slide window was mapped as an undirect graph. A graph is a pair of sets $G = (V, E)$, where V is a set of nodes (vertices, or points) so that each node represents the value of a test measurement for that day and E is a set of connections among the nodes of graphs. Therefore, each pair of nodes in a graph are connected by a link if there is a relationship between them [4, 5, 29]. The Euclidean distance has been widely used as a similarity measuring method [6, 18, 19]. Let $D_{ij} = 1, 2, 3, \ldots, M$ be the set of time series of M test measurements in each slide window. Each test measurement in a slide window is assigned to be a node in an undirected graph. Lets n_1 and n_2 be nodes in an undirected graph. They are connected if the distance (d) between them are less or equal to a determined threshold [13]:

$$(n_1, n_2) \in E, if \quad d(n_1, n_2) \leq \theta \tag{1}$$

where θ is a determined threshold. An example of an undirect graph is shown in Fig. 2. A graph G can be described by giving a square matrix $N \times N$ called adjacency or connectivity matrix A to describe the connections among the nodes of graph.
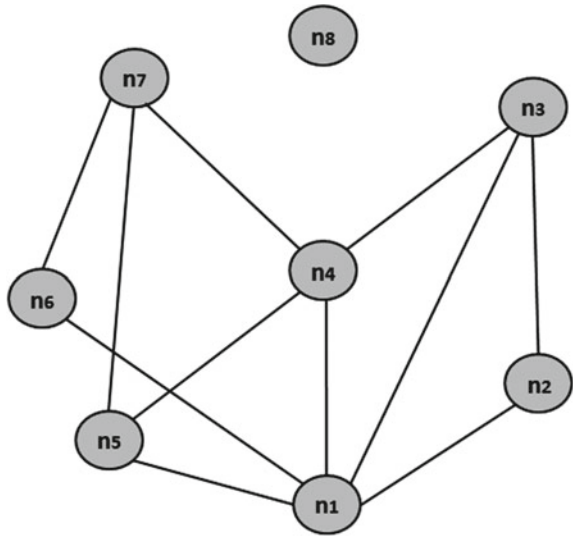
Fig. 2 An example of an
undirect graph

Table 1 The adjacency matrix of a graph G

	n_1	n_2	n_3	n_4	n_5	n_6	n_7
n_1	0	1	1	1	1	1	0
n_2	1	0	1	0	0	0	0
n_3	1	1	0	1	0	0	0
n_4	1	0	1	0	1	0	1
n_5	1	0	0	1	0	0	1
n_6	1	0	0	0	0	0	1
n_7	0	0	0	1	1	1	0

The adjacency matrix contains zeros in it's diagonal and thus it is a symmetric matrix.
The adjacency matrix is qual to one if there is a connect between two nodes, and zero
otherwise [6].

$$A(n_i, n_j) = \begin{cases} 1 & \text{if } (n_i, n_j) \in E, \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

For example, Table 1 shows the adjacency matrix of a graph G that consists from
7 nodes. We can notice that each element a_{ij} in an adjacency matrix A is equal to
1 when the connection exists, and zero otherwise. The diagonal of matrix A is still
zero for all it's elements.

2.3 Graph Features

The adjacency matrix of a graph G can be used to extract the statistical features of a graph G [13, 14, 26]. The statistical features of a graph can be used for prediction in this study. The following sections illustrate the most common extracted features from a graph G.

2.3.1 Degree Distributions of the Graph

The degree distribution, that denoted by $P(k)$, refers to the proportion of nodes with degree k divided by the total number of nodes in the graph [13]. It can be mathematically defined as follow:

$$P(k) = \frac{|\{n \mid d(n) = k\}|}{N} \tag{3}$$

where $d(n)$ refers to the degree of node n, N is the total number of nodes in the graph.

2.3.2 The Clustering Coefficient of the Graph

Clustering coefficient (CC) is one of the most important measures used to characterize the local and global structures of a graph [13, 26, 28]. Let n_i be a node in a graph G. Thus the local clustering coefficient of a given node n_i is computed as the proportion of links among n_i's neighbours which are actually realised compared with the total number of possible connections. For example, the clustering coefficient of a node n_3 in Fig. 2 is 1 because the node n_3 has three neighbours, which can have a maximum of 3 connections among them and all of them are realised. The overall level of clustering in a graph is measured as the average of the local clustering coefficients of all the nodes:

$$C' = \frac{1}{N} \sum_{i=1}^{N} C_{ni} \tag{4}$$

where, N is the number of nodes in a graph G and C_{ni} is the local clustering coefficient of the node n_i.

2.3.3 Jaccard Coefficient of the Graph

Jaccard Coefficient (it also called Jaccard Index) is a statistical tool that used to measure the similarity and diversity between two nodes of a graph [20]. Let n_i and n_j are two nodes in a graph G. Thus the Jaccard coefficient $\Gamma(n_i, n_j)$ is defined as the ratio

of the set of the neighboring intersections between those two nodes to the set of the neighboring unions for the two nodes. It can be mathematically defined as follows:

$$\Gamma(n_i, n_j) = \frac{|N(n_i) \cap N(n_j)|}{|N(n_i) \cup N(n_j)|} \quad (5)$$

where $N(n_i)$ is the set of neighbors of the node n_i that have an edge from n_i to them, and $N(n_j)$ is the set of neighbors of the node n_j that have an edge from n_j to them.

2.3.4   Average Degree

The average degree (AD) points out to the average number of links connecting in a node n_i to the other nodes in the graph [2]. The average degree of a graph can be defined as the total number of links for each node divided by the number of nodes in a graph [12]:

$$AD = \frac{1}{N} \sum_{i=1}^{m} K_i \quad (6)$$

where K_i is the degree of node n_i and N is the total number of nodes in a graph.

For example, we can easily calculate the degree of each node for a graph G shown in Fig. 2 and then calculate the average degree (AD) as follows:

$K(n_1) = 5, K(n_2) = 2, K(n_3) = 3, K(n_4) = 4, K(n_5) = 3, K(n_6) = 2, K(n_7) = 3,$ $K(n_8) = 0$, and $AD = 2$.

2.4   Bootstrap Aggregation (Bagging)

An ensemble approach is a very effective method that combines the decisions of multiple base classifiers in order to overcome the limited generalization performance of each base classifier and generate more accurate predictions than individual base classifier. Bootstrap aggregation, a.k.a bagging, is a machine learning ensemble algorithm designed to enhance the accuracy and stability of machine learning algorithms [15, 27]. In the bootstrap method, the classifiers are trained independently and then aggregated by an appropriate combination strategy. Specifically, our ensemble model can be divided into two phases. In the first phase, the model uses bootstrap sampling to generate a number of training sets. In the second phase, the training of the three base classifiers, i.e., Least Square-Support Vector Machine, Neural Network and Naive Bayes, is performed using the bootstrap training sets generated during the first phase. Figure 3 shows an example of the bagging algorithm which involves the three classifiers to build our ensemble model. In this study, the training set was divided into multiple datasets using the bootstrap aggregation approach, and then the classifiers were individually applied to these datasets to generate the final prediction.
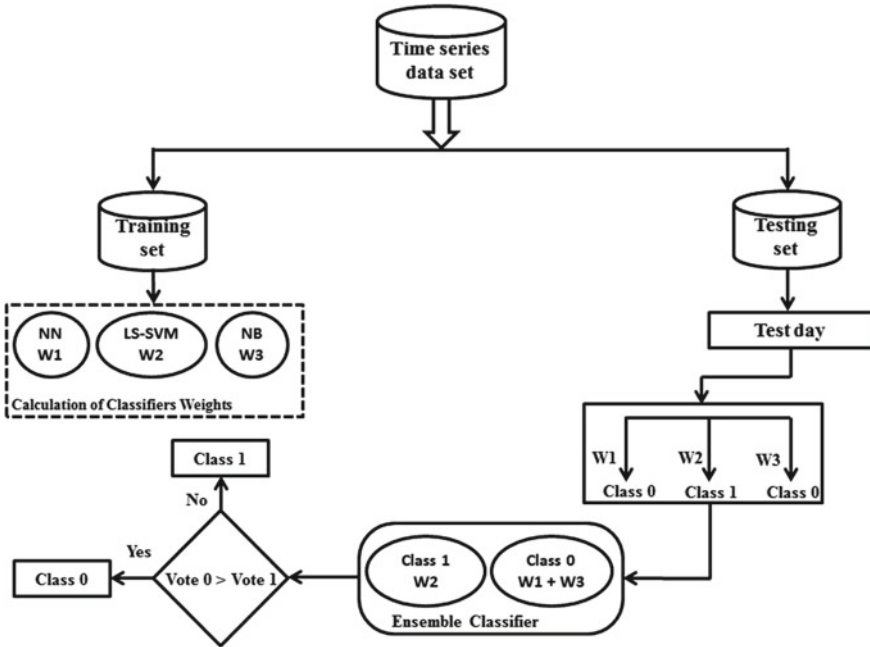
Fig. 3 An example of a bagging algorithm

It is noted that different individual classifier in the bagging approach may perform differently. Therefore, we assign a weight to each classifier's vote, based on how well the classifier performs. The classifier's weight is calculated based on its error rate. The classifier that has a lower error rate is considered more accurate and is therefore assigned a higher weight. The weight of classifier C_i's vote is calculated as follows:

$$w(C_i) = \log \frac{1 - error(C_i)}{error(C_i)}, 1 \leq C_i \leq 3 \tag{7}$$

The following example is presented to facilitate the understanding of our weighted bagging ensemble model:

1. Least Square-Support Vector Machine, Neural Network, and Naive Bayes are used as individual base classifier in the ensemble model. Suppose that the classifier training is performed on the training data and the error rate is calculated for each base classifier as 0.14 for LS-SVM, 0.25 for NN, and 0.30 for NB;
2. As per Eq. (7), the weight 0.78 is assigned to LS-SVM, 0.47 to NN, and 0.36 to NB;
3. Suppose that the three base classifiers generate the following predictions for a coming testing day: LS-SVM predicts 0, NN predicts 1, and NB predicts 1 (Here,

0 means no test is required on the testing day for a medical measurement; 1 means a test is required otherwise);

4. The ensemble classifier will use the weighted vote to generate the following prediction results:

   Class 1: NN + NB $\longrightarrow$ 0.47 + 0.36 $\longrightarrow$ 0.83,

   Class 0: LS-SVM $\longrightarrow$ 0.78.

5. Finally, according to the weighted vote, the class 1 has a higher value than class 0. Therefore, the ensemble classifier will classify this testing day as being in Class 1, suggesting that the patient in question need to take the test on that day for a medical measurement.

3 Experiment Result

This study aims at short-term risk assessment in chronic heart diseases patients based on analytic of a patient's historical medical data using structural graph similarity and machine learning-based ensemble classifier. As mentioned above, the time series slide windows were converted into undirect graphs. Then, the suitable features from graphs were extracted and entered as input features set for the ensemble classifier. The detailed experimental results are discussed in the following sub sections.

3.1 Performance Assessment

In this section, we present the details concerning the design of our experimental evaluation including datasets and performance metrics.

As the predictive performance of the proposed model is quite important, assessment of potential predictions is critically dependent on the quality of the used dataset. For this reason, telehealth data from Tunstall dataset will be conducted in this work. We use a real-life dataset obtained from our industry collaborator Tunstall to test the practical applicability of the proposed model. A Tunstall dataset obtained from a pilot study has been conducted on a group of heart failure patients and the resulting data were collected for their day-to-day medical readings of different measurements in a tele-health care environment. The Tunstall database employed in the development of the algorithm consists of data from six patients with a total of 7,147 different time series records. Data were acquired between May and January 2012, using a remote telehealth collaborator. The dataset is by nature in a time series and contains a set of measurements taken from the patients on different days. Each record in the dataset consists of a few different meta-data attributes about the patients such as patient-id, visit-id, measurement type, measurement unit, measurement value, measurement question, date and date-received. The characteristics of the features of the dataset are shown in Table 2.

Table 2 Characteristic features of the dataset

Feature name	Feature type
Id	Numeric
Id-patient	Numeric
HCN	Numeric
Visit-id	Numeric
Measurement type	Nominal
Measurement unit	Nominal
Measurement value	Numeric
Measurement question	Nominal
Date	Numeric
Date-received	Numeric

In addition, each record contains a few medical attributes including Ankles, Chest Pain, and Heart Rate, Diastolic Blood Pressure (DBP), Mean Arterial Pressure (MAP), Systolic Blood Pressure (SBP), Oxygen Saturation (SO$_2$), Blood Glucose, and Weight. Ethical clearance was obtained from the University of Southern Queensland (USQ) Human Research Ethics Committee (HREC) prior to the onset of the study. This dataset is used as the ground truth result to test the performance of our proposed model. The recommendations produced by our proposed model will be compared with the actual readings of the measurement in question recorded in the dataset to see how accurate our recommendations are.

Since a patient's historical medical data often have the class-imbalance problem (i.e., the number of normal data is much larger than that of the abnormal data), we carefully dealt with the class-imbalance problem when training the classifiers. The over-sampling and under-sampling methods have been used as good means to address this problem.

The selected input data were divided into two groups as the training and the testing sets. The slide windows time series data have been randomly divided into about 75% for the training of ensemble's classifier and 25% for the testing purpose. Several of experiments were designed and conducted to evaluate the proposed model using a real-life Tunstall database. Different sizes of slide windows were used to determine the best selected features set and the best size for each slide window as well. All the experimental results were conducted using MATLAB (R2015) on a desktop computer with the configurations of a 3.40 GHz Intel core i7 CPU processor with 8.00 GB RAM.

The performance of proposed method was evaluated by calculating the *accuracy*, *workload saving*, and *risk*. Accuracy refers to the percentage of correctly recommended days against the total number of days that recommendations are provided; workload saving refers to the percentage of the total number of days when recommendations are provided against the total number of days in the dataset, while risk refers to the percentage of incorrectly recommended days that recommendations are

no test needed. Mathematically, Accuracy, workload saving and risk are defined as follows [24]:

$$Accuracy = \frac{NN}{NN + NA} \times 100\% \tag{8}$$

$$Saving = \frac{NN + NA}{|\mathscr{D}|} \times 100\% \tag{9}$$

$$Risk = \frac{NR}{|\mathscr{D}|} \times 100\% \tag{10}$$

where *NN* denotes the number of days with correct recommendations, *NA* denotes the number of days with incorrect recommendations, *NR* denotes the number of days with incorrect days that recommendations are no test needed, and $|\mathscr{D}|$ refers to the total number of days in the dataset. Here, a correct recommendation means that the model produces the recommendation of "no test required" for the following day and the actual reading for that day in the dataset is normal. If this is a case, the recommendation is considered accurate.

3.2 Prediction Accuracy with Different Number of Features

We first carried out experiments to evaluate the recommendation performance of our system under different sets of statistical features extracted from the siding windows of the dataset. Several experiments are carried out to determine the best set of the graph features by which the original time series can be represented with the best form. The four graph features were tested separately to evaluate the prediction accuracy of the proposed system. Figure 4 shows the ranking of the statistical features based on their performance where the features were sorted in a descending order based on their effectiveness in predicting patient's condition.

3.2.1 Two-Features Set

To determine the best combination of the two graph features, a set of experiments was designed. In this experiment, at each time, a two features set of graphs was picked up from the ordered list in Fig. 4 and sent to ensemble classifier. The number of permutation of two graph features that was tested in this paper was six cases. Figure 5 shows the performance of the proposed method based on the graphs features. Based on the obtained results, it was observed that the combination of Jaccard coefficient and degree distribution recorded the highest accuracy of 81% compared to other combinations. We found that those two features were able to give the promising prediction. However, the lowest accuracy of 56% rate was recorded by the pair of
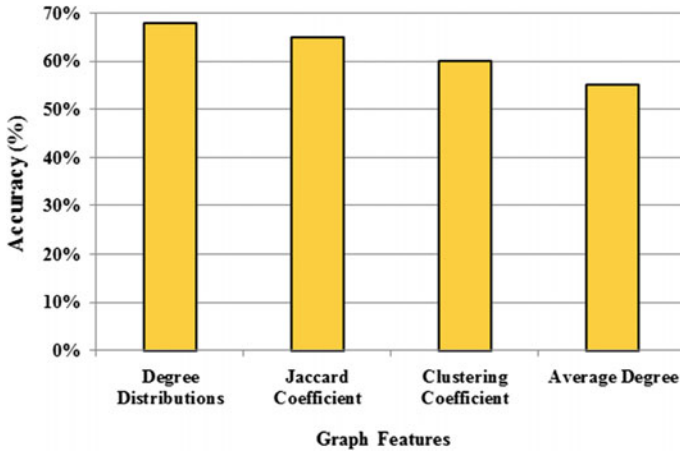
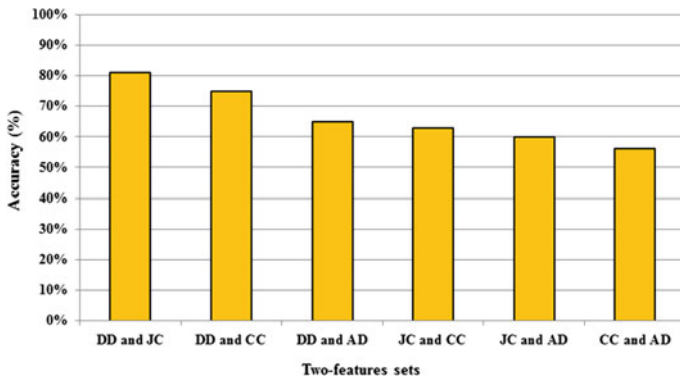**Fig. 4** Ranking of the graph features based on their accuracy performance

Fig. 5 Accuracy based on two-features sets (*Note DD* Degree Distribution; *JC* Jaccard Coefficient; *CC* Clustering Coefficient; *AD* Average Degree)

clustering coefficients and average degree. For further investigation, three features set was tested in the next experiment.

3.2.2 Three-Features Set

To assess the method ability to predict the status of patient with a high accuracy, the proposed method was tested using three features set. The first three graph features in Fig. 4 were selected. The three features were degree distribution, Jaccard coefficient and clustering coefficient. Figure 6 shows the performance of the proposed method using three and four features sets. The most noticeable results from this experiment were that the prediction accuracies were exceeded 94% compared with the sets of

Fig. 6 Accuracy based on three and four features sets
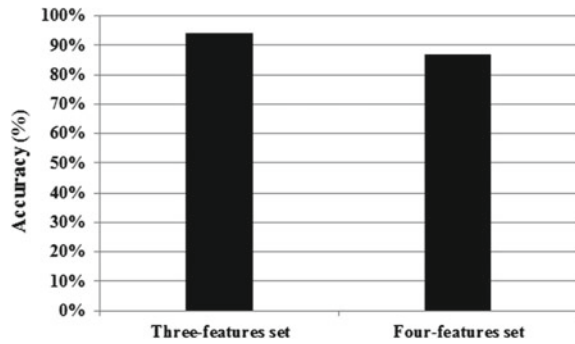
two features. For more accurate results, different experiments were designed with different data size. The results showed that there is a stability in the performance of the proposed method. Another three features set was also tested in this paper, however, the results confirmed that the three features set of degree distribution, Jaccard coefficient and clustering coefficient was the best combination of the graphs features to provide the recommendation accurately.

Four features was also tested and investigated in this paper. Based on the results in Fig. 6, the prediction accuracy of the proposed method was achieved a low rate compared with three features set. It was archived 87% using all the graph features including degree distribution, Jaccard coefficient, clustering coefficient and average degree. In this paper, the combination of the first three features for degree distribution, Jaccard coefficient and clustering coefficient was considered as they achieved the best accuracy.

3.3 Prediction Accuracy with Different Size of Slide Windows

The second influence in this work is associated to the size of window. In this experiment, the best window size is investigated to obtain the desired prediction accuracy. From the obtained results, it is clear that there is a positive relationship between the selected size of slide window and the predictive performance of the proposed system. It was found that when the number of nodes in a graph is increased due to the increasing the size of a slide window, the proposed method generates more accurate recommendations. To determine the optimum size of a slide, a set of experiments were conducted with different sizes of windows. It found that the model performance is improved by increasing the size of slide window (the number of nodes). This is because the characterises of time series data are clearly presented when the number of graph nodes is increased. Therefore, we tested our proposed model with different sizes of window and started with 7, 10, 15 and 20 days. In these experiments, the three features set of degree distribution, Jaccard coefficient and clustering coefficient were considered. The four Medical attributes including Heart Rate, Diastolic Blood

Table 3 Performance evaluation based on slide windows of 7 days

Measurement	Accuracy (%)	Saving (%)	Risk (%)
Heart rate	86.37	60.34	05.21
DBP	85.30	57.18	05.40
MAP	87.70	61.33	05.00
SO$_2$	84.30	55.44	05.90

Pressure (DBP), Mean Arterial Pressure (MAP), and Oxygen Saturation (SO$_2$) were used in the following experiments.

3.3.1 Slide Window of 7 Days

A slide windows of 7 days were used to test the predictive performance of the proposed method. Each day in the slide window was represented by a node in a graph. The three structural properties of the graphs were extracted and considered the key features to represent each window. The metrics of accuracy, workload saving and risk for all the graphs were calculated to verify the performance of proposed method. Table 3 presents the metrics of accuracy, workload saving and risk for each measurement in the Tunstall dataset.

Based on the obtained results, it was noticed that the performance of the proposed method was not good enough to predict the patient's condition due to the number of the graph nodes was not enough to reflect the behaviours of the time series data. To tackle this issue, the number of nodes in each graph was increased by considering a new size window. In the next experiment, the influence of using a window size of 10 days was discussed.

3.3.2 Slide Window of 10 Days

The time series data were segmented into windows by using a slide window of 10 days and then each window was transferred into a graph. As mentioned before, 10 days slide windows were considered to improve the accuracy of the proposed method and to make more accurate recommendations. One of the interesting findings in this paper, the proposed method yielded a high performance using 10 days slide window compared with the window size of 7 days. It can be noticed that the performance of proposed method significantly improved due to the number of graph nodes increased. It was found that the graphs nodes reflect big differences between the patient states which include whether he/she requires medical test or not. Table 4 shows the obtained results by the proposed method after considering the window size of 10 days.

Table 4 Performance evaluation based on slide windows of 10 days

Measurement	Accuracy (%)	Saving (%)	Risk (%)
Heart rate	92.75	59.80	03.95
DBP	91.40	58.77	04.50
MAP	90.55	60.65	04.80
SO$_2$	91.60	55.50	04.20

Table 5 Performance evaluation based on slide windows of 15 days

Measurement	Accuracy (%)	Saving (%)	Risk (%)
Heart rate	94.80	62.30	02.60
DBP	93.80	59.50	03.60
MAP	93.90	61.40	03.40
SO$_2$	94.60	61.80	02.90

Based on the obtained results in Table 4, It is interesting to note that the accuracies, for all the measurements, improved by more than 5% compared to the results in Table 3. In addition, using window size of 10 days did not considerably affect the performance of workload saving although the accuracy and risk are increased.

3.3.3 Slide Window of 15 Days

For further investigation, a window size of 15 days were adopted to test the performance of the proposed method. In this experiment, the size of window was increased into 15 day. Table 5 represents the metrics of accuracy, workload saving and risk for all the measurements using slide window size of 15 days. Based on results in Table 5, the average of accuracy of the proposed method were exceeded 94% across different measurements. The obtained results proved that the size of the window has a significantly potential on the accuracy of the prediction for all the measurements. One of the most important observations, the graphs characteristics were became significant to exhibit different behaviours when the patient state change from required test to not required test. We found that the connectivity among the graph nodes (clustering coefficients) are strong enough to reveal the difference between time series data.

Different sizes of window including 20, 25 and 30 days were also tested and evaluated in this study. It was noticed that there are no significant differences compared with the obtained results using the 15 days slide windows. Thus, the optimal window size was 15 days because it reflects the actual behavior of the time series data, on the basis of observation on the obtained results.

3.4 Comparative Study

To investigate the performance of our recommendation system, two performance comparisons were conducted in this section. In the first experiment, the performance of the recommendation system was evaluated based on individual classifier as well as machine learning-based ensemble classifier. In the second experiment, the proposed system was compared with some of our previous approaches. All the obtained results were recorded and evaluated.

3.4.1 Performance Evaluation Based on a Single Classifier as Well as Ensemble Model

In this experiment, we evaluate the performance of our system under 15 day slide windows and the three graph features set based on the previous information. Table 6 shows the results of comparison among the ensemble classifier and the individual classifiers. Based on the results, the system performance using individual classifier was between 80 and 85% across different measurements. The maximum accuracy of 85% was obtained by LS-SVM, while the minimum accuracy of 80% was gained by Nave Byes. We can notice that although the proposed system was conducted with different classifiers, there is no a big fluctuation in its performance and the accuracies of those classifiers are quite closer. One of the gold solution to improve the performance of the proposed method and to decrease the error rate is to combine multi-classifiers to classify the extracted features.

However, in this paper, an ensemble machine learning was used to classify the graphs features. Our recommendation system achieved a better prediction accuracy compared with the individual classifiers with an increase of 12%. As mentioned above, each classifier is trained and conducted with the dataset separately and then they combined according to an appropriate criteria. By comparing the results in Table 6, we can observe that the performance of the proposed sytem was escalated when the ensemble machine learning was adopted.

For more investigation, the execution time of the proposed model was calculated based on the ensemble classifier as well as individual classifiers. Figures 7 and 8 show the complexity time for each individual classifier and the ensemble model. we observed that the ensemble model takes more time to complete the training and

Table 6 Performance evaluation based on a single classifier and an ensemble model

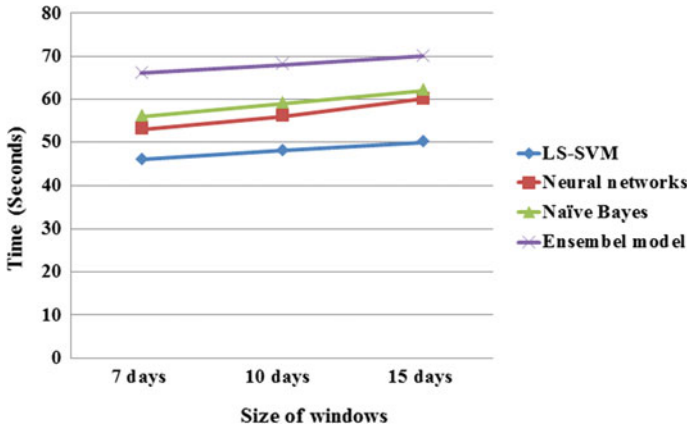| Classifier	Accuracy (%)	Saving (%)	Risk (%)
LS-SVM	85.30	61.10	04.30
Neural network	83.50	60.80	05.10
Naive bayes	80.40	60.95	04.90
Ensemble model	94.27	61.25	03.01

Fig. 7 Comparison of the execution time between the classifiers and the ensemble model under different slide windows
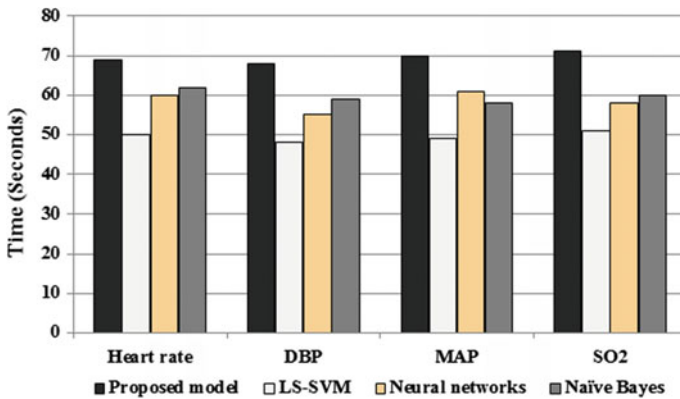

Fig. 8 Comparison of the execution time between the classifiers and the ensemble model under different measurements

prediction than the individual base classifier. This is reasonable as the ensemble model needs to aggregate the results from the base classifiers to generate the weights for them and produce the final recommendation. The ensemble model sacrifices a little on the execution time for achieving better recommendation effectiveness for patients. Additionally, the training stage can be performed off-line so that it will not adversely affect the efficiency in generating recommendations for patients during the prediction stage.

Table 7 Prediction accuracy comparison with other methods

Tunstall dataset

Method	Size of window	Techniques used	Accuracy (%)
Raid et al. [23]	5 days	Basic heuristic algorithm	86
Raid et al. [24]	5 days	Basic heuristic algorithm, Regression-based algorithm and Hybrid algorithm	91
Proposed method	15 day	Structural graph similarity and ensemble model	**94**

3.4.2 Effectiveness Comparison with Previous Approaches

To evaluate the performance of the proposed method, the prediction results were compared with some of our previously proposed methods that tackle the exactly same problem as we do in this paper using the same Tunstall dataset for a fair comparison. Table 7 represents the performances comparison among the two other reported methods and our proposed method. Based on results, the proposed model is the best among the three methods. Raid et al. [23] used a innovative time series prediction algorithm to provide recommendations to heart disease patients in the tele-health environment. The best accuracy was achieved using slide windows of 5 days. The average of the accuracy for all patients they achieved was 86% across all measurements. An intelligent recommender system, supported by three innovative predictive algorithms, was proposed by Raid et al. [24] for short-term risk assessment on patients in telehealth environment. The size of slide window was empirically detected by 5 days as the best accuracy in this study. The average of accuracy results obtained was 91% for all measurements. It clearly seems from the above results in Table 7 that the proposed model yielded the highest accuracy compared with the two others methods using the same dataset.

4 Conclusions and Future Research Directions

In this work, we propose a recommendation system supported by the structural graph properties and advanced machine learning ensemble for short-term disease risk prediction and medical test recommendation in the telehealth environment for patients suffering from chronic heart disease. This study applies the structural graph, which effectively represents the medical time series data and input the extracted statistical features to the ensemble model to generate the accurate, reliable recommendations

for chronic heart disease patients. Three popular and capable classifiers, i.e., Least Square-Support Vector Machine, Neural Network, and Naive Bayes are used to construct the ensemble framework.

The experimental results show that the proposed system using slide windows of 15 day with the optimal statistical features set produced by the structural graph properties yields a better predictive performance for all measurements. The results also show that our system using the ensemble classifier with optimal features set can correctly predict up to 94% of the subjects across all measurements. It is also observed that our system is more effective than the individual base classifiers used in the ensemble model and outperforms the previously proposed approaches to solve the same problem. Our evaluation establishes that our recommendation system is effective in improving the quality of clinical evidence-based decisions and help reduce the time costs incurred by the chronic heart disease patients in taking their daily medical test, whereby improving their overall life quality.

There are several directions for our future research work in this study. First, we want to evaluate our proposed system using additional appropriate datasets which preferably have a large number of data records. We are also interested in applying other ensemble techniques, such as boosting and Adaboost, to produce recommendations and conducting a comparative study on those different ensemble models. Finally, given the generality of our proposed model in dealing with medical time series data, we will explore the possibility to apply our system to support telehealth care for patients suffering from other type of diseases.

References

1. Abegunde, D. O., Mathers, C. D., Adam, T., Ortegon, M., & Strong, K. (2007). The burden and costs of chronic diseases in low-income and middle-income countries. *The Lancet*, *370*, 1929–1938.
2. Artameeyanant, P., Sultornsanee, S., & Chamnongthai, K. (2015). Classification of electromyogram using weight visibility algorithm with multilayer perceptron neural network. In: *7th International Conference on Knowledge and Smart Technology (KST)* (pp. 190–194). Chonburi, Thailand: IEEE.
3. Bashir, S., Qamar, U., & Khan, F. H. (2015). BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. *Australasian Physical and Engineering Sciences in Medicine*, *38*, 305–323.
4. Bernhardt, B. C., Bonilha, L., & Gross, D. W. (2015). Network analysis for a network disorder: The emerging role of graph theory in the study of epilepsy. *Epilepsy and Behavior*, *50*, 162–170.
5. Blondel, V. D., Gajardo, A., Heymans, M., Senellart, P., & Van Dooren, P. (2004). A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM Review*, *46*, 647–666.
6. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*, 175–308.
7. Braamse, A. M., Jean, C. Y., Visser, O. J., Heymans, M. W., van Meijel, B., Dekker, J., et al. (2016). Developing a risk prediction model for long-term physical and psychological functioning after hematopoietic cell transplantation. *Biology of Blood and Marrow Transplantation*, *22*, 549–556.

8. Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123–140.

9. Chang, C. D., Wang, C. C., & Jiang, B. C. (2011). Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. *Expert Systems with Applications*, *38*, 5507–5513.

10. Chen, D., Jin, D., Goh, T. T., Li, N., & Wei, L. (2016). Context-awareness based personalized recommendation of anti-hypertension drugs. *Journal of Medical Systems*, *40*, 202.

11. Dewar, A. R., Bull, T. P., Malvey, D. M., & Szalma, J. L. (2017). Developing a measure of engagement with telehealth systems: The mHealth technology engagement index. *Journal of Telemedicine and Telecare*, *23*, 248–255.

12. Diykh, M., & Li, Y. (2016). Complex networks approach for EEG signal sleep stages classification. *Expert Systems with Applications*, *63*, 241–248.

13. Diykh, M., Li, Y., & Wen, P. (2016). EEG sleep stages classification based on time domain features and structural graph similarity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *24*, 1159–1168.

14. Fang, Z., & Wang, J. (2014). Efficient identifications of structural similarities for graphs. *Journal of Combinatorial Optimization*, *27*, 209–220.

15. Gao, H., Jian, S., Peng, Y., & Liu, X. (2016). A subspace ensemble framework for classification with high dimensional missing data. *Multidimensional Systems and Signal Processing*, 1–16.

16. He, Y., & Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Current Opinion in Neurology*, *23*, 341–350.

17. Huang, F., Wang, S., & Chan, C.C. (2012). Predicting disease by using data mining based on healthcare information system. In: *Granular Computing (GrC)*, (pp. 191–194). Hangzhou, China: IEEE.

18. Huang, X., & Lai, W. (2006). Clustering graphs for visualization via node similarities. *Journal of Visual Languages and Computing*, *17*, 225–253.

19. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, *31*, 264–323.

20. Kogge, P. M. (2016). Jaccard Coefficients as a Potential Graph Benchmark. In *Parallel and Distributed Processing Symposium Workshops* (pp. 921–928) Chicago: IEEE.

21. Krishnaiah, V., Narsimha, D. G., & Chandra, D. N. S. (2013). Diagnosis of lung cancer prediction system using data mining classification techniques. *International Journal of Computer Science and Information Technologies*, *4*, 39–45.

22. Kuh, D., & Shlomo, Y. B. (2004). *A Life Course Approach to Chronic Disease Epidemiology*. Oxford University Press.

23. Lafta, R., Zhang, J., Tao, X., Li, Y., & Tseng, V. S. (2015). An intelligent recommender system based on short-term risk prediction for heart disease patients. In *Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 102–105). Singapore: IEEE.

24. Lafta, R., Zhang, J., Tao, X., Li, Y., Tseng, V. S., Luo, Y., et al. (2016). An intelligent recommender system based on predictive analysis in telehealthcare environment. *Web Intelligence*, *4*, 325–336.

25. Lang, S. (2017). Cognitive eloquence in neurosurgery: Insight from graph theoretical analysis of complex brain networks. *Medical Hypotheses*, *98*, 49–56.

26. Li, X., Hu, X., Jin, C., Han, J., Liu, T., Guo, L., et al. (2013). A comparative study of theoretical graph models for characterizing structural networks of human brain. *International journal of biomedical imaging*, 2013.

27. Li, S., Tang, B., & He, H. (2016). An imbalanced learning based MDR-TB early warning system. *Journal of Medical Systems*, *40*, 164.

28. Micheloyannis, S., Pachou, E., Stam, C. J., Vourkas, M., Erimaki, S., & Tsirka, V. (2006). Using graph theoretical analysis of multi channel EEG to evaluate the neural efficiency hypothesis. *Neuroscience Letters*, *402*, 273–277.

29. Miraglia, F., Vecchio, F., & Rossini, P. M. (2017). Searching for signs of aging and dementia in EEG through network analysis. *Behavioural Brain Research*, *317*, 292–300.

30. Mohktar, M. S., Redmond, S. J., Antoniades, N. C., Rochford, P. D., Pretto, J. J., Basilakis, J., et al. (2015). Predicting the risk of exacerbation in patients with chronic obstructive pulmonary

disease using home telehealth measurement data. *Artificial Intelligence in Medicine*, *63*, 51–59.

31. Njie, G. J., Proia, K. K., Thota, A. B., Finnie, R. K., Hopkins, D. P., Banks, S. M., et al. (2015). Clinical decision support systems and prevention: A community guide cardiovascular disease systematic review. *American Journal of Preventive Medicine*, *49*, 784–795.

32. Panzica, F., Varotto, G., Rotondi, F., Spreafico, R., & Franceschetti, S. (2013). Identification of the epileptogenic zone from stereo-EEG signals: A connectivity-graph theory approach. *Frontiers in Neurology*, 4.

33. Polat, K., & Gne, S. (2007). Breast cancer diagnosis using least square support vector machine. *Digital Signal Processing*, *17*, 694–701.

34. Sarsoh, J. T., Hashem, & K. M. (2012). Classifying of human face images based on the graph theory concepts. *Global Journal of Computer Science and Technology*, 12.

35. Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review*, *1*, 27–64.

36. Snchez, A. S., Iglesias-Rodrguez, F. J., Fernndez, P. R., & de Cos, Juez F. (2016). Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *International Journal of Industrial Ergonomics*, *52*, 92–99.

37. Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics*, *1*, 3.

38. Thong, N. T. (2015). HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. *Expert Systems with Applications*, *42*, 3682–3701.

39. Tuffry, S. (2011). *Data Mining and Statistics For Decision Making*, Wiley Chichester.

40. Valentini, G., Masulli, F. (2002). Ensembles of learning machines. In *Italian Workshop on Neural Nets* (pp. 3–20). Heidelberg: Springer.

41. Vecchio, F., Miraglia, F., Piludu, F., Granata, G., Romanello, R., Caulo, M., et al. (2017). Small World architecture in brain connectivity and hippocampal volume in Alzheimers disease: A study via graph theory from EEG data. *Brain Imaging and Behavior*, *11*, 473–485.

42. Verma, L., Srivastava, S., & Negi, P. (2016). A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. *Journal of Medical Systems*, *40*, 1–7.

43. Vural, C., & Yildiz, M. (2010). Determination of sleep stage separation ability of features extracted from EEG signals using principle component analysis. *Journal of Medical Systems*, *34*, 83–89.

44. Wang, J., Qiu, S., Xu, Y., Liu, Z., Wen, X., Hu, X., et al. (2014). Graph theoretical analysis reveals disrupted topological properties of whole brain functional networks in temporal lobe epilepsy. *Clinical Neurophysiology*, *125*, 1744–1756.

45. Wang, J., Qiu, M., & Guo, B. (2017). Enabling real-time information service on telehealth system over cloud-based big data platform. *Journal of Systems Architecture*, *72*, 69–79.

46. Yang, J. G., Kim, J. K., Kang, U. G., & Lee, Y. H. (2014). Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFISLDA). *Personal and Ubiquitous Computing*, *18*, 1351–1362.

47. Yeh, D. Y., Cheng, C. H., & Chen, Y. W. (2011). A predictive model for cerebrovascular disease using data mining. *Expert Systems with Applications*, *38*, 8970–8977.

48. Zhang, J., Lafta, R., Tao, X., Li, Y., Zhu, X., Luo, Y., et al. (2017). Coupling a fast fourier transformation with a machine learning ensemble model to support recommendations for heart disease patients in a telehealth environment. *5*, 10674–10685.