CrossMark

A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases

Subhashini Narayan[1] · E. Sathiyamoorthy[1]

Abstract

Recently, using of the intelligent technologies in the field of clinical decision making is increased rapidly to improve the lifestyles of patients and to help for reducing the workload and cost concerned in their healthcare. Heart diseases are one of the primary causes of death. However, if the diseases are identified at the early stage, the rate of death can be decreased. Thus, the disease identification process has become a matter of concern. An efficient medical recommendation system has been proposed in this paper, namely Fourier transformation-based heart disease prediction system (FTHDPS) by using Fourier transformation and machine learning technique to predict the chronic heart diseases effectively. Here, the input sequences rely on the patient's time series details or data, which are crumbled by Fourier transformation for extracting the frequency information. In FTHDPS, a bagging model is utilized for predicting the conditions of the patients in advance to produce the absolute recommendation. In FTHDPS, three classifiers are used, namely artificial neural network, Naive Bayes and support vector machine, and real-life time series chronic heart disease data are used to evaluate the proposed model. The experimental results demonstrate that FTHDPS is much efficient to provide a reliable and accurate recommendation to the heart patients.

Keywords Intelligent system · Naive Bayes classifier · Neural network · Support vector machine · Rough sets · Heart disease

1 Introduction

Day-by-day heart diseases are rapidly increasing, and gradually, it has become the major reason of death in all over the world [1]. Some heart diseases are: heart failure, myocardial infarction, heart arrhythmia, cardiomyopathy, angina pectoris, atrial fibrillation, etc. Heart diseases mainly affect the functions and structure of the heart, and the diseases have many features. Doctors and medical practitioners usually used these features to detect the heart diseases, and sometimes, it is difficult for them to diagnosis the diseases accurately and quickly. So, a computer-based model can be highly beneficial to diagnosis the heart diseases at the early stage, so that the causes of these diseases can be decreased up to an extent.

In today's intelligent word, it is very difficult to use and store huge amount of medical data, and cloud computing and big data are widely used to store these huge amount of medical data [2–13]. It can be meaningless, if there is no advanced technology to use or extract information from these huge amount of medical data [14]. Data mining techniques have significant impact over the decades to extract the information from a dataset and to predict the human's behavior [15]. These techniques are useful for converting any stored data into some valuable information that supports decision support ability, i.e., to enhance the accuracy and speed of a diagnosis system. Recently, disease prediction system (DPS) and data mining techniques have a great attention [16]. One of the most popular data mining tools is single-layer perceptron (SLP) classifier that is used in many applications to predict the diseases [15]. Some other popular techniques are Naive Bayes classifier,

✉ Subhashini Narayan
  subanarayan@gmail.com

[1] School of Information Technology and Engineering, VIT University, Vellore, India

SVM and so on. Naive Bayes classifier is based on Bayes' Theorem. In Naive Bayes classifier, it is assumed that the existence of any particular feature set in any class is not related to other feature. For an example, any fruit can be considered as apple, if the color is red and round. Even when these features are dependant on each other, all these properties contribute independently to the probability that the particular fruit is an apple. This model is very simple and efficient also for a very large datasets. Along with its simplicity, it is fast to predict the class for any test dataset and also performs efficiently for multi-class prediction. SVM is a supervised learning model that analyzes data or information for classification as well as for regression. It is mainly based on decision planes. A decision plane separates a number of objects, which have different class memberships. SVM algorithm can be implemented by using a kernel for performing linear classification. SVM is very efficient for real-world and complex problems like hand-writing recognition, image and text classification, etc. It can also perform for the datasets having many attributes.

Many researchers have proposed many advanced technologies to predict the heart diseases by using many important features or attributes of the heart, such as blood pressure, gender, fasting cholesterol, blood sugar, etc. [17–22]. However, in most of the datasets, there are some relevant as well as some irrelevant features along with the redundant features. Redundant features are unable to improve the results. However, creates some noises, which may degrade the results or the system may run very slowly. Therefore, removing of those redundant features is mandatory prior to apply any classification technique. Sometimes, too many features in a datasets may slow the execution. Thus, reduction in the irrelevant features or attributes is also mandatory for an efficient heart disease diagnosis system. In 2013, Shilaskar and Ghatol [23] have proposed a diagnosis system for predicting the heart disease. This scheme consists of classification and feature selection techniques. Here, SVM classifier is combined with the inclusion of forward feature. However, this technique is quite complex. Shao et al. [24] have proposed a novel technique for identifying the heart diseases at the early stage. Their scheme is mainly based on the reduction of the features or attributes. After reduction in the features, the remaining attributes are used to classify by using artificial neural network classifier. However, this scheme is slow. Guan et al. [25] have introduced a novel feature selection scheme for predicting the heart diseases. Here, SVM classifier is used to predict the disease at the early state. Their experiments give 76.5% accuracy, when 12 features or attributes are selected out of 44. Nowadays, rough set theory is used for reducing the number of subset from a dataset. Rough sets can be applied for removing the redundant attributes and to select a number of significant

attributes that can produce high prediction accuracy and high speed. By using greedy heuristics algorithms and rough set, many researchers have proposed many diagnosis systems for predicting the heart diseases [26]. To find the best minimal attributes, another solution is used to combine the rough set with metaheuristic algorithms [27]. Firefly algorithm (FA) can be considered as one of the swarm intelligent techniques, which is a metaheuristic approach or algorithm that depends on the behavior of the fireflies of the nature. FA algorithm is more efficient than the genetic algorithms and practical swarm optimization technique to find the global optima [28]. These schemes have high accuracy than the traditional schemes, and these schemes can be also applied for optimization problems, NP hard problems, etc. All these existing schemes are helpful to predict the heart diseases. However, these existing schemes are slow to predict the heart diseases. Moreover, accuracy is not high, which is one of the major criteria for any heart disease prediction system.

A novel scheme has been introduced here in this paper for an efficient heart disease prediction, which can minimize the aforementioned problems. The proposed diagnostic system has been developed by using fast Fourier transformation and machine learning approaches for providing accurate recommendations that is necessary to do a medical test in the future. This recommendation system is mainly based on the heart conditions of the patients, i.e., time series data or information of past days. The performance of this novel proposed recommendation system is evaluated in terms of accuracy, risk and workload saving. The experiments are executed by using time series data, which are gathered from a number of heart disease patients, and the results are compared with other classifiers, namely Naive Bayes, neural network, etc.

The main contributions of this paper are mentioned below:

- A novel system has been introduced in the paper on the basis of fast Fourier transformation and machine learning approach to predict the heart diseases at the early stage.
- FTHDPS is evaluated by using real-life time series information or data. The proposed recommendation system is highly helpful for the medical practitioners to diagnosis the heart diseases.
- Experimental results demonstrate the proficiency or efficiency of the proposed diagnosis system over the existing models.

The remaining part of this paper is structured into different parts: related works are presented in Sect. 2. In Sect. 3, background studies have been discussed. The proposed heart disease diagnosis system is presented in Sect. 4. Results and discussions have been presented in

Sect. 5. Finally, conclusions of this paper are given in Sect. 6.

2 Related works

Tsumoto [29] has introduced a novel approach in 2000 for medical data. Here, in this paper, Tsumoto has presented that data mining techniques can be considered very useful to find hidden patterns from any database or dataset that may be reused for stored dataset because human begins are not able to maintain a huge number of dataset having numerous records. Characteristic of the clinical data is also discussed in this paper by Tsumoto.

In 2003, Neagoe et al. [30] have proposed a model based on neuro-fuzzy approach for diagnosing heart diseases. This scheme is mainly distinguished the electrocardiogram (ECG) signals by using two steps: removal of features from the ECG signals and pattern classification. This proposed scheme is experimented by using a database of ECG containing 40 subjects in which 20 subjects are having heart diseases and rest of the subjects are normal ones. However, accuracy of this scheme is not high to predict the diseases, so it might not be trustable for the doctors and medical practitioners.

Ordonez [31] has proposed a rule for identifying and predicting the heart diseases at the early stage. Here, attributes of the risk factors are also considered for the heart diseases, and three processes are proposed by Ordonez to reduce the number of patterns. The experimental results illustrate that the rule is very efficient to decrease the chances of heart diseases. However, the presence of heart diseases is anticipated in this scheme, which may not be a suitable solution in many cases.

In 2006, Noh et al. [32] have proposed an associative classification for diagnosis the cardiovascular disease. They have extracted multi-parametric features or attributes by using the preprocessing technique. The proficient frequent pattern (FP) growth technique is the fundamental of this method. Their rule of cohesion plays an important role on pruning patterns. The several rules and pruning and dataset consist of 670 participants, which are grouped into two categories, namely patients having coronary artery disease and normal people. However, accuracy is not high in this scheme.

Koutsojannis and Hatzilygeroudis [33] have proposed a novel approach by using neuro-fuzzy technique for medical science. This paper represents the using of fuzzy Adaline neurons and differential evolution to optimize the membership functions. Based on this scheme, many schemes have been proposed by many researchers to predict the heart diseases. However, their scheme is slow at the time of diagnosis the diseases.

In 2008, Tsipouras et al. [34] have introduced a novel method by using fuzzy logic and data mining approach for diagnosing coronary artery disease. Here, a fuzzy logic rule-based decision system is used in this paper to diagnosis the diseases. There are four stages in this proposed scheme, namely induction of the decision tree, extraction of rules, alteration of the crisp set into a fuzzy logic model and optimization. For experiments, 199 subjects are used in this decision model, which show significant results for diagnosis the heart diseases at the early stage.

Vazirani et al. [35] have proposed a novel model for identifying the heart diseases by using modular neural network. They have mainly discussed about two ways for diagnosis the heart diseases, namely manual and automatic diagnosis. In the automatic diagnosis system, modular neural network and intelligent expert system are used for diagnosis the diseases of heart. In this system, attributes are divided based on some properties, which are given into two different modules for training and testing, namely radial basis function neural network (RBFNN) and backpropagation neural network (BPNN). However, this scheme is slow.

Anooj [36] has introduced a fuzzy rule-based decision support system for the heart diseases by using computer-aided technologies. The automatic processes to diagnosis the diseases is one of the main advantages of this scheme. This proposed scheme consists of two steps for diagnosis the heart diseases, namely weighted fuzzy rules generation phase and fuzzy rule-based decision system development phase. In this scheme, suitable attributes or features are generated by using data mining procedure, and then, these features are used for generating the fuzzy rules. Then, these rules are weighted on the basis of occurrence in the datasets. However, this scheme is not much efficient to predict the heart diseases.

Shao et al. [24] have proposed a novel technique to for identifying the heart diseases at the early stage. Their scheme uses multivariate adaptive regression splines, logistic regression and rough set approaches to reduce the features or attributes for predicting the heart diseases. After reduction in the features, the remaining attributes are used to classify by using artificial neural network classifier. The datasets are taken from UCI datasets for training and testing the model. However, logistic regression method of this scheme only reduces one feature and it helps to increase 2% accuracy. Moreover, this scheme is slow.

3 Background studies

A detailed description has been presented in this section regarding background studies. The proposed model consists of mainly three machine learning classifiers, namely

Naive Bayes (NB), least square SVM (LSSVM) and artificial neural networks (ANN).

3.1 Artificial neural network

ANN is considered as a supervised learning technique, which can be utilized for providing an efficient solution for numerous complex problems. It uses numerous number of processing nodes or neurons that are interconnected to each other. This network is sometimes considered as the directed graph, where each and every neuron i performs the transfer function f_i by using below Eq. (1) [37]:

$$y_i = f_i \left(\sum_{j-1}^{n} w_{ij} x_j - \theta_i \right) \qquad (1)$$

where output of the neuron i is represented by y_i, input of the jth neuron is represented by x_j, and the weight of connection between two neurons i and j is represented by w_{ij}. Threshold bias is represented by θ, and f_i is used for taking nonlinear, such as a sigmoid, Gaussian function, Heaviside, etc.

3.2 Least square support vector machine

LSSVM [38] is also considered as a supervised learning technique, where statistical learning theory is used. It is mainly used in medical science for predicting heart diseases [39], breast cancer, etc. [40].

A linear LSSVM is mainly designed for classifying a dataset, which keeps two different classes represented by using $\{1, -1\}$. LSSVM attempts to set the data containing high-dimensional space. It then utilizes a hyperplane separation technique to maximize the distance between the support vectors and plane. For an example if a training dataset holds n data denoted by $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m) \in R^m$, LSSVM classifier selects optimal separating hyperplane for separating the classes. This classifier works based on the following rules to find a solution of the given problem:

$$y_i[(wx_i) + w_0] = 1 - \varepsilon_i, \quad \text{where} \quad i = 1, 2, \ldots, m \qquad (2)$$

$$\frac{1}{2}||w||^2 + \frac{c}{2}\sum_{i=1}^{m} \varepsilon_i^2 \qquad (3)$$

Equations (2) and (3) show the problem, which is solved by using LSSVM, and it can be shown by Eq. (4).

$$
\begin{aligned}
(w, b, \alpha, \varepsilon) = &\frac{1}{2}||w||^2 + \frac{c}{2}\sum_{i=1}^{m} \varepsilon_i^2 \\
&- \sum_{i=1}^{m} \alpha_i \{y_i[(wx_i) + w_0] - 1 + \varepsilon_i\}
\end{aligned} \qquad (4)
$$

3.3 Naive Bayes

Naive Bayes algorithm is mainly used for creating models that have the predictive capabilities. Naive Bayes classifier provides the new approaches to understand and explore data, and it is mainly used, when the inputs have high dimension [41]. Here, in the proposed technique, Naive Bayes classifier takes the characteristics or attributes of the patients having heart diseases. This classifier is executed in the below cases:

- When the dataset have high dimension.
- When the attributes or characteristics are not dependent on each other.
- Exhibits the speed and high accuracy, when the classifier is applied to large databases.

According to the Bayes theorem, the probability of a problem can be denoted by using Eq. (5).

$$P(C_x|Y) = \frac{P(Y|C_x)P(C_x)}{P(Y)} \qquad (5)$$

where Y is a dataset that needs to classify. Class label is denoted by C_x, and the probability of belongingness of Y to C_x is denoted as $P(C_x|Y)$.

4 Proposed scheme

The proposed scheme has been designed by using fast Fourier transformation and machine learning technique to provide medical recommendations to the users or patients, who are suffering from the heart diseases. In this section, FTHDPS has been discusses in details.

In FTHDPS, the medical test-related data of a patient are segmented into the sliding windows, which are used during data analysis or data processing. Each and every sliding window has been passed throughout fast Fourier transformation for extracting the statistical attributes or features. Fast Fourier transformation is used to study about the features of the medical data, which are difficult to attain in the time domain. The extracted features, which are from the data, are taken into the proposed learning model as the input for making a decision-making system, which can predict the heart diseases of a patient on the coming days. Figure 1 shows the architecture of the proposed recommendation system.

4.1 Fast Fourier transformation (time series data)

Fast Fourier transformation (FFT) can be referred as an efficient approach for calculating discrete Fourier transform (DFT), and FFT is also useful for computing the
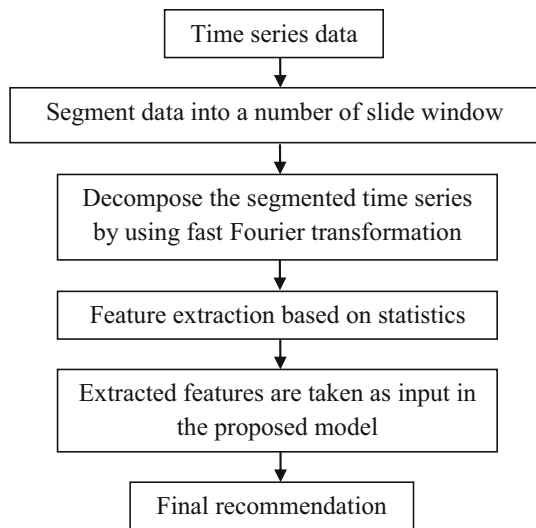
Fig. 1 Architecture of the proposed recommendation system

inverse of DFT. The data sequence of input is decomposed by using DFT for extracting the frequency information to predict the condition of the patient in advance.

The input is scaled by using any normalization technique prior to perform FFT. The variables of normalization technique fall within the range of $[0, 1]$ [42]. Equation (6) is used for performing normalization.

$$D_{norm} = \frac{D_{orig} - D_{min}}{D_{max} - D_{min}} \tag{6}$$

where normalized data and original raw data are represented by using D_{norm} and D_{orig}, respectively. The maximum and minimum data values are represented by using D_{max} and D_{min}, respectively.

Now, let $x(t)$ be the time series for a sliding window. DFT of $x(t)$ can be shown by using Eq. (7).

$$X(c^{jw}) = \sum_{t=-\infty}^{\infty} x(t)c^{-jwt} \tag{7}$$

where the discrete time index and the frequency are represented by using t and w, respectively. The transform pair regarding DFT is defined by Eq. (8), when T number of time series are presented.

$$X(P) = \sum_{t=0}^{T-1} x(t)W_T^{tp} \Leftrightarrow x(t) = \frac{1}{T}\sum_{p=0}^{T-1} X(P)W_T^{-tp} \tag{8}$$

where $W = c^{-\frac{j2\Pi}{T}}$.

In addition, DFT can be defined as the discrete time Fourier transformation by using the following equations:

$$x = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(T-1) \end{bmatrix}, \quad X = \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(T-1) \end{bmatrix} \tag{9}$$

$$W = W_T^{tp} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & W_T & W_T^{T-1} \\ \vdots & \vdots & \vdots \\ 1 & W_T^{T-1} & W_T^{(T-1)(T-1)} \end{bmatrix} \tag{10}$$

The association between X and x is represented by Eq. (11).

$$X = Wx \Leftrightarrow x = \frac{1}{T}W^H X \tag{11}$$

For a given input signal $x(t)$ consists of length T, DFT matrix W needs T^2 multiplications on the basis of the above equation. Thus, the requisite implementation cost to factorize FFT into a matrix denoted by W is less than the cost of the simple Fourier transformation because each and every phase of FFT requires only T additions and $\frac{T}{2}$ multiplications [43, 44].

In the proposed system, the input data are denoted as $X = \{y_1, y_2, \ldots, y_n\}$ that contains n data, and it is segmented by using a number of sub-segments on the basis of predefined value (k), which specifies sliding window's size regarding each and every sub-segment. Here, data are analyzed by using FFT for extracting frequency details for predicting the condition of the patient. In FTHDPS, few frequency bands denoted as $\alpha, \beta, \gamma, \delta$, and θ have been obtained by using FFT for each and every sliding window. High-frequency band is divided into eight parts because it captures almost all the information or details of a sliding window. As a reference, the original or main sliding window has been added with the extracted features set. In total, 14 frequency bands have been generated as a result for each and every slide window. Moreover, FFT coefficient's power has been evaluated for all the 14 bands. Thus, there are $14 + 14 = 28$ frequency bands, which are extracted from each and every sliding window.

Eight separate statistical features or attributes can be extracted from each and every frequency band. The extracted attributes or features have been denoted by using $X_{Max}, X_{Min}, X_{SD}, X_{Med}, X_{Mean}$, XFQ, XRG and XSQ. The dataset keeps dependent features give best performance. In FTHDPS, the extracted features or attributes from each and every band are clustered into a single vector, and then, it is taken for the proposed model as the input to predict the condition of the patient or user.

4.2 Bootstrap aggregation or bagging

The proposed model is very efficient, which combines the results or decisions of several classifiers to improve the performance of each and every base classifier by providing accurate predictions. Bootstrap aggregation or bagging is one of the machine learning algorithms designed for enhancing the stability and accuracy [45, 46]. It can be used for both regression and classification, and always used by using decision tree. Here, multiple models are used, and then, aggregate the results of all the models for final prediction. All these models must be specialized in a specific part of the feature space, which support accurate predictions and maximum use. There are mainly two parts in bootstrap aggregation, namely bootstrapping and aggregation. Bootstrapping can be considered as sampling technique, and in aggregation, all the results from different models are aggregated. The proposed model is mainly divided into two steps. Here, at first, the system model utilizes bootstrap sampling technique for generating many training sets. Then, in the second step, three base classifiers, namely NB, neural network and LSSVM, are performed on the training sets, which are generated in the first step. Here, the training set is separated into the multiple datasets by applying the bootstrap aggregation technique. The above-mentioned classifiers are then applied individually on datasets for generating the concluding prediction. Thus, in the bagging approach, different classifiers can perform individually or differently. So, a weight is assigned for each and every classifier on the basis of how all these classifiers are actually performed, and this weight is calculated on the basis of the error rate. The classifier, which holds less error rate, has been treated more accurate, and then, assigned high weight. The classifier's weight C_i is calculated by using Eq. (12).

$$w(C_i) = \log \frac{1 - \text{error}(C_i)}{\text{error}(C_i)}, \quad 1 \leq C_i \leq 3 \qquad (12)$$

The weighted bagging technique is presented in the following example:

1. Neural network (NN), LSSVM and Naive Bayes classifiers are used individually in the proposed technique. Let on the training data the classifiers are performed, and for each classifier, the error rate is taken as 0.25 for NN, 0.30 for NB and 0.14 for LSSVM.
2. By using Eq. (12), weight 0.78 has been assigned for LSSVM, 0.47 for NN and 0.36 for NB.
3. Let the classifiers provide the below predictions, which is mainly for the coming testing day:

   a. LSSVM predicts 1.
   b. NN predicts 0.
   c. NB predicts 0.

   In the above terminology, 0 implies no test is needed for a medical recommendation and 1 implies a test is needed.

4. The proposed technique uses the weight for generating the below prediction recommendation:

   a. Class 0: NB + NN = 0.36 + 0.47 = 0.83.
   b. Class 1: LSSVM = 0.78.

5. At last, as per the weight, class 1 has lesser value than class 0. Thus, the proposed classifier classifies the test as being in Class 0.

5 Results and discussion

Experiments are executed for evaluating the efficiency of FTHDPS. In this section, the details of results are presented.

5.1 Experimental setup

Real-life datasets are used to evaluate FTHDPS. Here, Tunstall datasets are gathered from a pilot research, which are performed on numerous heart disease patients or users. The collected details or data hold day to day details of patients regarding different medical measurements. The dataset is considered mainly a time series data, which holds six patient's details. There are in total 8148 time series details. Each and every record or detail in the dataset holds several metadata attributes of the patient, namely visit identity (ID), patient ID, measurement question, measurement value, date, measurement unit, date received and measurement type. Table 1 shows the characteristics regarding the attributes of the metadata. Several medical measurements of many patients of each day for a specific time period are there in the dataset, namely heart rate, mean arterial pressure (MAP), diastolic blood pressure

Table 1 Metadata attributes

Attribute name	Attribute type
Visit ID	Numeric
Patient ID	Numeric
Measurement question	Nominal
Measurement value	Numeric
Date	Numeric
Measurement unit	Nominal
Date received	Numeric
Measurement type	Nominal

(DBP), blood glucose, weight and oxygen saturation (SO$_2$), where heart rate, SO$_2$, MAP and DBP are taken for the evaluation purpose.

The datasets are separated into two different parts for the evaluation purpose, namely testing set and training set. Three base classifiers of the proposed model or system are trained by utilizing the training dataset, and then, it is validated. In these experiments, 25% datasets are used as the testing data and the remaining 75% are used as the training dataset. The recommendations performed for each measurement of medical tests by the proposed model are compared with the authentic value of the testing dataset for evaluating the efficiency of the system.

The performance of all three classifiers and the proposed model have been evaluated based on three metrics, namely workload saving, accuracy and risk. Workload saving can be referred as the percentage of the amount of days at the time of recommendations providing to skip the medical examinations [49] with respect to the amount of days of the training dataset. Accuracy can be referred as the percentage of the accurate recommended days with respect to the amount of days. Risk can be referred as the percentage of wrong recommended days with respect to the amount of days of the training dataset. Workload saving, accuracy and risk can be defined by the following equations:

$$\text{Workload saving} = \frac{NN + NA}{|D|} \times 100\% \qquad (13)$$

$$\text{Accuracy} = \frac{NN}{NN + NA} \times 100\% \qquad (14)$$

$$\text{Risk} = \frac{NR}{|D|} \times 100\% \qquad (15)$$

where NN = Number of days for which accurate or correct recommendations are given. NA = Number of days for which wrong recommendations are given. NR = Number of days for which risky recommendations are provided. |D| = Number of days of the training dataset.

Here, an accurate recommendation implies that the system recommends for no test required or test required. By using MATLAB, the proposed recommendation system has been developed on a computer system of Intel core i5 processor, 3.40 GHz, 8 GB RAM and 1 TB hard disk.

5.2 Effectiveness of the recommendation for different statistical features

For appraising the performance of the novel proposed recommendation system, many statistical features are extracted from the dataset.

At first, eight statistical features or attributes are tested differently for evaluating the proposed model's accuracy. Figure 2 shows the statistical features ranking on the basis
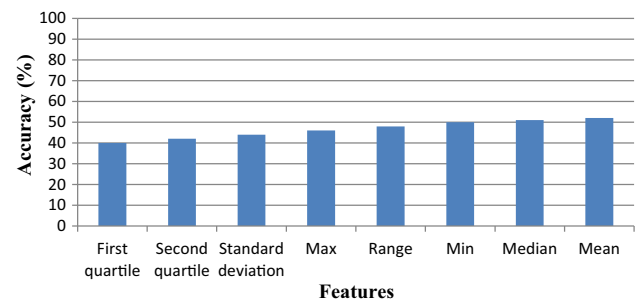
Fig. 2 Accuracy of the statistical features

of the performance. Here, features are placed in the ascending order on the basis of the effectiveness to predict the condition of the patient. The extracted statistical attributes or features are taken as the input in this experiment to train the base classifiers.

With respect to the different number of attributes or features, the proposed recommendation system tested for its effectiveness. The performance of the proposed system for different features ranges starting from two to eight are experimented and recorded. Table 2 shows the experimental results. It can be observed from the results that a positive correlation is there between the extracted statistical features or attributes and the proposed system's performance regarding workload saving, accuracy and risk. Thus, the proposed system is much better, when the total amount of the features are increased.

## 5.3 Comparisons of the effectiveness with the existing schemes

The experiments presented in the previous subsection shown that the proposed system is better regarding the performance of the recommendation. The proposed system is compared in this subsection with some other existing schemes. A heuristic method is proposed in [47] based on heuristic rules for the patients of heart disease supported by a telehealth system. In [48], a novel recommendation system has been proposed by Lafta et al. by combining

Table 2 Performance of FTHDPS for different number of features

Feature sets	Workload saving (%)	Accuracy (%)	Risk (%)
2	66.32	85.31	7.00
3	66.71	86.21	6.23
4	66.86	89.56	5.94
5	68.10	90.12	4.56
6	68.23	91.85	4.18
7	70.42	92.79	3.24
8	71.80	93.56	2.75

Table 3 Comparisons of the recommendation performance with the existing scheme

Method	Workload saving (%)	Accuracy (%)	Risk (%)
[47]	10	86	8
[48]	15	91	5
Proposed	71	93	2
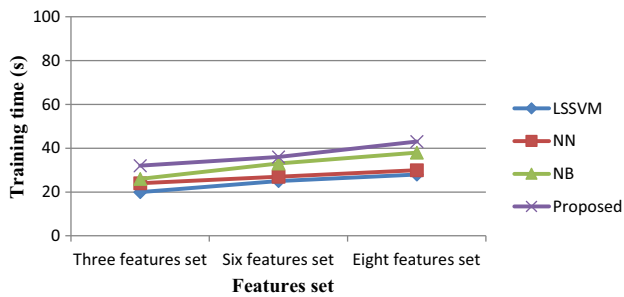
Fig. 3 Training time comparisons
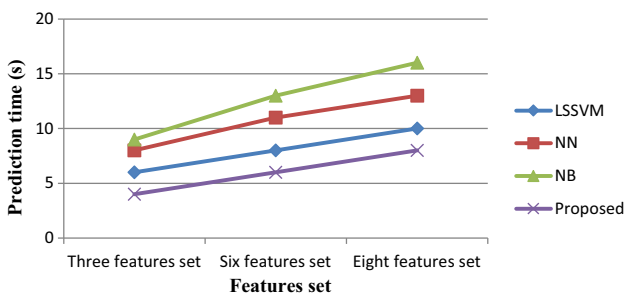
Fig. 4 Prediction time comparisons

heuristic approach, regression-based prediction technique and hybrid approach [50].

Table 3 represents the comparison results of effectiveness. From Table 3, it can be easily identified that the proposed model or scheme is much better than the other schemes since it provides 93% accuracy. FTHDPS saves workload of 71%, where the risk is also less compared to the other existing schemes. Thus, FTHDPS is effective than the existing schemes.

5.4 Efficiency comparisons with the existing schemes

To evaluate the efficiency, the execution time has been calculated for different datasets for the proposed scheme as well as for all the classifiers. Figures 3 and 4 show the calculated execution time for the proposed scheme and for the classifiers in the prediction and training stages.

From Figs. 3 and 4, the following statement can be made:

1. For all the models, the training stage has more execution time than the prediction stage.
2. For different feature sets, linear execution time for the proposed model and for all three classifiers supports the efficiency to provide fast recommendation.
3. The proposed model consumes less time among all the classifiers for completing the prediction and training. Thus, FTHDPS can provide the recommendation in less time to the patients, which can be very effective for the doctors or medical practitioners.

6 Conclusions

Heart diseases are one of the major issues of death. In this paper, a novel recommendation system has been proposed (FTHDPS) on the basis of fast Fourier transformation technique and machine learning approach to predict the heart diseases at the early stage. FTHDPS uses the patient's time series data or details for the past days. The performance of FTHDPS is evaluated in terms of workload saving, accuracy and risk. Real-life time series data are used for the experimental purpose. The results of the experiments prove that the proposed model provides high accuracy as well as minimize the workload for the patients. Moreover, FTHDPS also achieves low risk for wrong recommendations. In future, FTHDPS can be evaluated for the cancer diseases to predict it at the early stage.

Compliance with ethical standards

Conflict of interest There is no conflict of interest between the authors to publish this manuscript.

References

1. Pu LN, Zhao Z, Zhang YT (2012) Investigation on heart risk prediction using genetic information. IEEE Trans Inf Technol Biomed 16(5):795–808
2. Namasudra S (2017) An improved attribute-based encryption technique towards the data security in cloud computing. Concurr Comput Pract Exerc 1:5. https://doi.org/10.1002/cpe.4364
3. Namasudra S, Roy P, Balamurugan B (2017) Cloud computing: fundamentals and research issues. In: Proceedings of the 2nd international conference on recent trends and challenges in computational models, IEEE, Tindivanam, India
4. Namasudra S, Nath S, Majumder A (2014) Profile based access control model in cloud computing environment. In: Proceedings of the international conference on green computing, communication and electrical engineering, IEEE, Coimbatore, India, pp. 1–5

5. Namasudra S, Roy P (2017) A new secure authentication scheme for cloud computing environment. Concurr Comput Pract Exerc 29(20):e3864. https://doi.org/10.1002/cpe.3864

6. Namasudra S, Roy P (2016) Secure and efficient data access control in cloud computing environment: a survey. Multiagent Grid Syst Int J 12(2):69–90

7. Namasudra S, Roy P (2017) Time saving protocol for data accessing in cloud computing. IET Commun 11(10):1558–1565

8. Namasudra S, Roy P, Balamurugan B, Vijayakumar P (2017) Data accessing based on the popularity value for cloud computing. In: Proceedings of the international conference on innovations in information, embedded and communications systems (ICIIECS), IEEE, Coimbatore, India

9. Namasudra S, Roy P (2015) Size based access control model in cloud computing. In: Proceedings of the international conference on electrical, electronics, signals, communication and optimization, IEEE, Visakhapatnam, India, pp. 1–4

10. Namasudra S, Roy P (2017) A new table based protocol for data accessing in cloud computing. J Inf Sci Eng 33(3):585–609

11. Namasudra S, Roy P (2018) PpBAC: popularity based access control model for cloud computing. J Org End User Comput 30(4):14–31

12. Sarkar S, Saha K, Namasudra S, Roy P (2015) An efficient and time saving web service based android application. SSRG Int J Comput Sci Eng 2(8):18–21

13. Namasudra S (2018) Cloud computing: a new era. J Fundam Appl Sci 10(2):113–135

14. Li M, Yu S, Ren K, Lou W (2010) Securing personal health records in cloud computing: patient-centric and fine-grained data access control in multi-owner settings. In: Proceedings of the international conference on security and privacy in communication systems, pp. 89–106

15. Liu X, Lu R, Ma J, Chen L, Qin B (2016) Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification. IEEE J Biomed Health Inform 20(2):655–668

16. Mathew G, Obradovic Z (2011) A privacy-preserving framework for distributed clinical decision support. In: Proceedings of the computational advances in bio and medical sciences, pp. 129–134

17. Polat K, Güneş S (2007) Breast cancer diagnosis using least square support vector machine. Digit Signal Proc 17(4):694–701

18. Sánchez AS, Iglesias-Rodríguez FJ, Fernández PR, Juez FJDC (2016) Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. Int J Ind Ergon 52:92–99

19. Huang F, Wang S, Chan CC (2012) Predicting disease by using data mining based on healthcare information system. In: Proceedings of the IEEE international conference on granular computing, pp. 191–194

20. Krishnaiah V, Narsimha DG, Chandra NS (2013) Diagnosis of lung cancer prediction system using data mining classification techniques. Int J Comput Sci Inf Technol 4(1):39–45

21. Das R, Turkoglu I, Sengur A (2009) Effective diagnosis of heart disease through neural networks ensembles. Expert Syst Appl 36(4):7675–7680

22. Bashir S, Qamar U, Khan FH (2015) BagMOOV: a novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. Australas Phys Eng Sci Med 38(2):305–323

23. Shilaskar S, Ghatol A (2013) Feature selection for medical diagnosis: evaluation for cardiovascular diseases. Expert Syst Appl 40(10):4146–4153

24. Shao YE, Hou CD, Chiu CC (2014) Hybrid intelligent modeling schemes for heart disease classification. Appl Soft Comput 14(5):47–52

25. Guan W, Gray A, Leyffer S (2009) Mixed-integer support vector machine. In: Proceedings of the NIPS workshop on optimization for machine learning, pp. 1–6

26. Hoa NS (1996) Some efficient algorithms for rough set methods. In: Proceedings IPMU'96 Granada, Spain, pp. 1541–1547

27. Ye D, Chen Z, Ma S (2013) A novel and better fitness evaluation for rough set based minimum attribute reduction problem. Inf Sci 222:413–423

28. Yang XS (2009) Firefly algorithms for multimodal optimization. In: Proceedings of the 5th international conference on stochastic algorithms: foundations and applications. Springer, Berlin, pp. 169–178

29. Tsumoto S (2000) Problems with mining medical data. In: Proceedings of the 24th annual international computer software and applications conference, IEEE, Taipei, Taiwan

30. Neagoe VE, Iatan IF, Grunwald S (2003) A neuro-fuzzy approach to classification of ECG signals for ischemic heart disease diagnosis. In: The proceedings of the AMIA Annual Symposium, pp. 494–498

31. Ordonez C. (2004). Improving heart disease prediction using constrained association rules. In: Seminar presentation at University of Tokyo

32. Noh K, Lee HG, Shon HS, Lee BJ, Ryu KH (2006) Associative classification approach for diagnosing cardiovascular disease (LNCIS, 345). Springer, Berlin, pp 721–727

33. Koutsojannis C, Hatzilygeroudis I (2007) Using a neurofuzzy approach in medical application (LNCS, 4693). Springer, Berlin, pp 477–484

34. Tsipouras MG, Exarchos TP, Fotiadis DI, Kotsiam AP, Vakalis KV, Naka KK, Michalis LK (2008) Automated diagnosis of coronary artery disease based on data mining and Fuzzy modeling. IEEE Trans Inf Technol Biomed 12(4):447–458

35. Vazirani H, Kala R, Shukla A, Tiwari R (2010) Use of modular neural network for heart disease. Int J Comput Commun Technol 1(2–4):88–93

36. Anooj PK (2012) Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J King Saud Univ Comput Inf Sci 24(1):27–40

37. Yao X (1999) Evolving artificial neural networks. Proc IEEE 87(9):1423–1447

38. Suykens JAK, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300

39. Bai Y, Han X, Chen T, Yu H (2015) Quadratic kernel-free least squares support vector machine for target diseases classification. J Comb Opt 30(4):850–870

40. Sharawardi NSA, Choo YH, Chong SH, Muda AK, Goh OS (2014) Single channel sEMG muscle fatigue prediction: an implementation using least square support vector machine. In: Proceedings of the 4th world congress on information and communication technologies, IEEE, Bandar Hilir, Malaysia, pp. 320–325

41. Han J, Kamber M, Pei J (2011) Data mining concepts and techniques, 3rd edn. Morgan Kaufmann Publishers, Burlington

42. Singh YN, Gupta P (2007) Quantitative evaluation of normalization techniques of matching scores in multimodal biometric systems (LNCS, 4642). Springer, Berlin, pp 574–583

43. Brigham EO (1988) The fast Fourier transform and its applications. Prentice-Hall, Englewood Cliffs

44. Alfred M (1999) Signal analysis wavelets, filter banks, time-frequency transforms and applications. Wiley, New York

45. Li S, Tang B, He H (2016) An imbalanced learning based MDRTB early warning system. J Med Syst 40(7):1–9

46. Gao H, Jian S, Peng Y, Liu X (2016) A subspace ensemble framework for classification with high dimensional missing data. Multidimens Syst Signal Process 28(4):1309–1324

47. Lafta R, Zhang J, Tao X, Li Y, Tseng VS (2015) An intelligent recommender system based on short-term risk prediction for heart disease patients. In: Proceedings IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT), IEEE, Singapore, Singapore, pp. 102–105

48. Lafta R, Zhang J, Tao X, Li Y, Tseng VS, Luo Y, Chen F (2016) An intelligent recommender system based on predictive analysis in telehealthcare environment. Web Intell 14(4):325–336

49. Rizwan P, Rajsekhara Babu M, Suresh K (2017) Design and development of low investment smart hospital using internet of things through innovative approaches. Biomed Res 28(11):4979–4985

50. Rizwan P, Babu MR, Balamurugan B, Suresh K (2018) Real-time big data computing for internet of things and cyber physical system aided medical devices for better healthcare. In: Majan international conference (MIC), IEEE, pp. 1–8