Machine learning-based classification of cardiac diseases from PCG recorded heart sounds

Anjali Yadav[1] · Anushikha Singh[2] · Malay Kishore Dutta[1] · Carlos M. Travieso[3]

Abstract

Cardiovascular diseases are one of the most fatal diseases across the globe. Clinically, conventional stethoscope is used to check the medical condition of a human heart. Only a trained medical professional can understand and interpret the heart auscultations clinically. This paper presents a machine learning-based automatic classification system based on heart sounds to diagnose cardiac disorders. The proposed framework involves strategic processing and framing of heart sound to extract discriminatory features for machine learning. The most prominent features are selected and used to train a supervised classifier for automatic detection of cardiac diseases. The biological abnormalities disturbing the physical functioning of the heart cause variations in the auscultations, which is strategically used in terms of some discriminatory features for machine learning-based automatic classification. The proposed method achieved 97.78% accuracy with the equal error rate of 2.22% for abnormal and normal heart sound classification. The experimental results exhibit that the performance of the proposed method in proper diagnosis of the cardiac diseases is high in terms of accuracy and has low error rate which makes the proposed algorithm suitable for real-time applications.

Keywords Body auscultation · Cardiac disease · Feature extraction · p value · Machine learning · Automatic classification

1 Introduction

Cardiac diseases are one of the major causes of mortality worldwide. According to World Health Organization (WHO), cardiac diseases are leading cause of death in the world which is about 17.3 million deaths per year [1]. In the developing countries, this situation gets worse due to the lack of medical professionals. However, diagnosis of cardiac diseases at early stages can make the treatment economical, efficient and helpful in preventing death. Heart sound examination is the most common and basic diagnostic technique used by doctors to examine the heart health. Heart sounds are generated by flow of blood through it and beating of heart. Clinically, body auscultations are of crucial importance in determining health of human beings. Auscultation is the act of listening body sounds produced by heart, lungs, blood vessels or other organs using electrical stethoscope [2]. In cardiac auscultation, a doctor may use his stethoscope to hear distinct sounds that provide important data about heart's condition. A vibration is generated due the pressure of blood flow, opening or closing of heart valve as well as due to the contraction of cardiac muscle. This vibration propagates up through the tissues to the thorax, and it is the method to measure the heart sound. The presence of the problem in heart is identified through murmurs. These murmurs are identified as abnormal heart sound. The turbulent blood flow in the heart system is the reason for murmurs. It is very important to check the pitch and timing of the sounds for diagnosis of the heart condition.

✉ Malay Kishore Dutta
mkd@cas.res.in

Anjali Yadav
anjali@cas.res.in

Anushikha Singh
Anushikha.Singh@dbst.iitd.ac.in

Carlos M. Travieso
carlos.travieso@ulpgc.es

1 Centre for Advanced Studies, Dr. APJ Abdul Kalam Technical University, Lucknow, India

2 Indian Institute of Technology, Delhi, India

3 Signals and Communication Department, IDeTIC, University of Las Palmas de Gran Canaria, Las Palmas, Spain

Medically heart auscultation is an essential part for examination of heart to diagnose cardiac diseases at the initial stage. Stethoscope is used for the examination of human heart auscultations. It is an easy, efficient way as well as a computationally cheap technique, but a trained medical professional is needed to understand and interpret heart sounds [3]. Machine learning-based normal and abnormal heart sound classification using audio processing may be a possible method for automatic diagnosis of cardiac diseases or abnormalities without the help of trained medical professionals. Automatic diagnosis of cardiac diseases using heart auscultations maybe of great help in primary health centers for early diagnosis and screening of cardiac disorders.

With reference to audio processing-based diagnosis of heart diseases, some relevant work is summarized in this section. Azmy [4] proposed an algorithm for abnormal and normal classification of heart sounds using new mother wavelet and support vector machines (SVM). The coefficients of new mother wavelet transform are extracted and then these components are ordered using SVM. It is done in two stages, first is discrete wavelet transform which is separated from new mother wavelet transform and then the factual segments are ascertained utilizing SVM. Mayorga et al. [5] had classified heart and lung sounds based on events. Maglogiannisaet.al. [6] proposed an algorithm using support vector machine (SVM) to classify heart valve disease. Gogineni et al. [7] developed a model for heart risk stratification using support vector machine (SVM) and learning vector quantization (LVQ), among which SVM gave 99% accuracy for the prediction of classes with normal, first stroke, second stroke and end of life of patients. Fu et al. [8] had proposed an algorithm for heart sound diagnosis using DTW (dynamic time warping) and MFCC (Mel-frequency cepstral coefficients) on the heart signals. Gudadhe et al. [9] proposed an effective decision support system for cardiac diseases based on MLP neural system technique and SVM. Ghumbre et al. [10] reported an algorithm for heart disease diagnosis on the basis of the support vector machine (SVM) using sequential minimal optimization and for the radial basis function (RBF), orthogonal least square calculation is used. SVM gave the highest accuracy (89.43%) and RBF gave the least (79.30%). Sheela et al. [11] proposed a framework using cardiovascular expectation which is used to predict the sudden cardiac death with the help of the time and frequency of heart rate variability from SVM and ECG. Sonawane et al. [12] had proposed an algorithm for the coronary illness utilizing learning vector quantization (LVQ) neural network calculation. The algorithm is divided into two levels, in the first step, the attributes are treated as the input in the algorithm and then in the next step, LVQ is used to train the algorithms. Pedreira et al.

[13] proposed a model on data input selection for the coronary illness.

Table 1 shows a comparative chart of some more existing work in this field of normal and abnormal heart sound classification for diagnosis of cardiac diseases. The dataset used in these various studies are local datasets and not same. The comparison of the different methods should be done on the same database for any scientific merit. Since we have no access to these local datasets, this task remains difficult.

Still there is a requirement to design a more accurate and efficient system for automated diagnosis of heart diseases from heart sounds which can work in real time. Automated classification of normal and abnormal heart sound maybe helpful in early diagnosis of cardiac diseases.

The main contribution of this paper is an automatic machine learning-based system for classification of the phonocardiogram (PCG) recorded heart auscultations between normal and abnormal heartbeat. Bandpass filter is applied to the recorded heartbeat signals to remove noises and other redundant signals which are recorded by the PCG device along with the heartbeat of the patient. The usage of the filter helps in increasing the credibility of the algorithm and give a better accuracy rate in classification of the phonocardiogram (PCG) recorded heart sound signals.

Another contribution of the paper is to divide each heart sound sample into non-overlapping frames and then extract features from them. This will help in improving the accuracy of the developed algorithm. After extracting the features from each frame of the recorded signals, statistical features such as mean and standard deviation were calculated from all frames as a single representative of audio sample. The audio samples are strategically processed to extract informative discriminatory features to determine abnormal and normal heart sounds using supervised classification.

For better classification of the developed algorithm, most prominent features having lower p-value determined from the statistical models of significance were only subjected to classifiers. Another significant contribution of the proposed work is its low time complexity for normal and abnormal heart sound classification. Considering only most discriminatory features for classification improves the performance of classifiers and ensures less time complexity and makes it acceptable for real-time applications.

The remaining structure of this paper work is as follows: Section 2 provides information about the database which has been used to test the developed algorithm. Section 3 describes about the proposed audio signal processing-based methodology for proper classification of the abnormal and normal heart sounds. It includes all the steps, like preprocessing of PCG recordings of heart sounds, feature analysis and selection of discriminatory features using p-values and

Table 1 Comparative table indicating the performance comparison of existing works with the proposed method

Reference	Objective	Data base used	Feature considered	Criteria for feature selection	Classification method	Accuracy	Equal error rate (EER)
Vadicherla et al. [14]	Decision support system for heart diseases based on sequential minimal optimization in support vector machine	Local heart database 297 recordings	Sequential minimal optimization (SMO)	Not reported	SVM	90.57%	Not reported
Kumar et al. [15]	Heart murmur classification with feature selection	Local data base 81 recordings	17 features extracted	SFSS	SVM	95.74%	Not reported
Guraksin et al. [16]	Classification of heart sound based on the least squares support vector machine	Local database of 120 recordings	Wavelet features	Not reported	LS-SVM	Not reported	Not reported
Hadi et al. [17]	Classification of heart sounds on the basis of wavelets and neural networks	Local data base 250 recordings	Wavelet coefficients	Not reported	Multilayer perceptron neural network	92%	Not reported
Singh et al. [18]	Heart sounds classification using feature extraction of phonocardiography signal	PeterJ Bentley 60 recordings	23 features from frequency domain, time domain, cepstrum domain and statistical features	Ranker and Info gain attribute evaluation	Bayes Net, Naïve Bayes, Logit Boost, SGD	93.33%	Not reported
Mandal et al. [19]	A new framework for wavelet-based analysis of acoustical cardiac signal	Local database	Wavelet basis	Additive criterion proposed by Coifman	Sure shrink	Not reported	Not reported
Salleh et al. [20]	Classification of heart sounds based on multipoint auscultation system	Centre for biomedical engineering	MFCC and HMM	SVD	Not reported	94.80%	Not reported
Proposed work	Automated classification of normal and abnormal PCG recorded heart sounds	National Institute of Health (National Center for Research Resources)	12 statistical features	Wilcoxon rank sum rule (p-value $\leq 10–14$)	SVM, k-NN, random forest, Naïve Byes	97.78%	2.22%

training of different machine learning algorithms. Section 4 describes the experimental results achieved by the proposed algorithm. Section 4 gives an account about the experimental results obtained from the proposed algorithm and finally, in the last Sect. 5, a discussion about the conclusion of the proposed work is mentioned.

2 Database used: heart sounds—materials

In the proposed methodology, experiments are accomplished on the PCG recorded heart sound signals database which is created by National Institute of Health (National Center for Research Resources) [21]. The PCG recorded heart signals are either from healthy subjects or

pathological patients. Each audio belongs to one patient only, and the existence of same patient file is most unlikely in the sets. This database contains both normal as well as abnormal sounds for training and testing purpose.

The heart sounds were recorded from different body locations including four main locations: Aortic area, mitral area, tricuspid area and pulmonic area. These four locations are mostly used by doctors to listen heart sounds. The database is basically categorized into two parts: abnormal and normal heart sound. The abnormal heart sounds were recorded from patient who had a confirmed cardiac disease, typically related to defects of heart valve and coronary artery while the normal sounds were recorded from healthy subjects. Recorded heart sounds of this database belong to both, children as well as adults. All the recorded heart

signals are sampled at a frequency of 2000 Hz and are accessible in.wav format. Figure 1 shows two samples of normal/abnormal heart sound recording and their power spectrum. It is clearly observed from the power spectrum diagram that the abnormal heart sound signal has more frequency content as compare to normal heart sound.

3 Proposed methodology: methods

The main aim of the proposed work is to develop a machine learning-based system which can be used for automatic diagnosis of normal and abnormal heart sound

using PCG recorded heart beats. The proposed method is explained in Fig. 2 shows the flowchart diagram of the proposed framework for automated diagnosis of cardiac diseases. The proposed framework involves strategic processing and framing of heart sounds to extract informative features followed by an analysis to determine the most prominent features for classification. All extracted features are analyzed based on their p-values, and only those features are subjected to different classifiers which are discriminatory in nature. The main purpose to select the discriminatory features is to obtain a more accurate and precise classification of normal/abnormal heart sounds. The proposed model is categorized into four sub-sections:
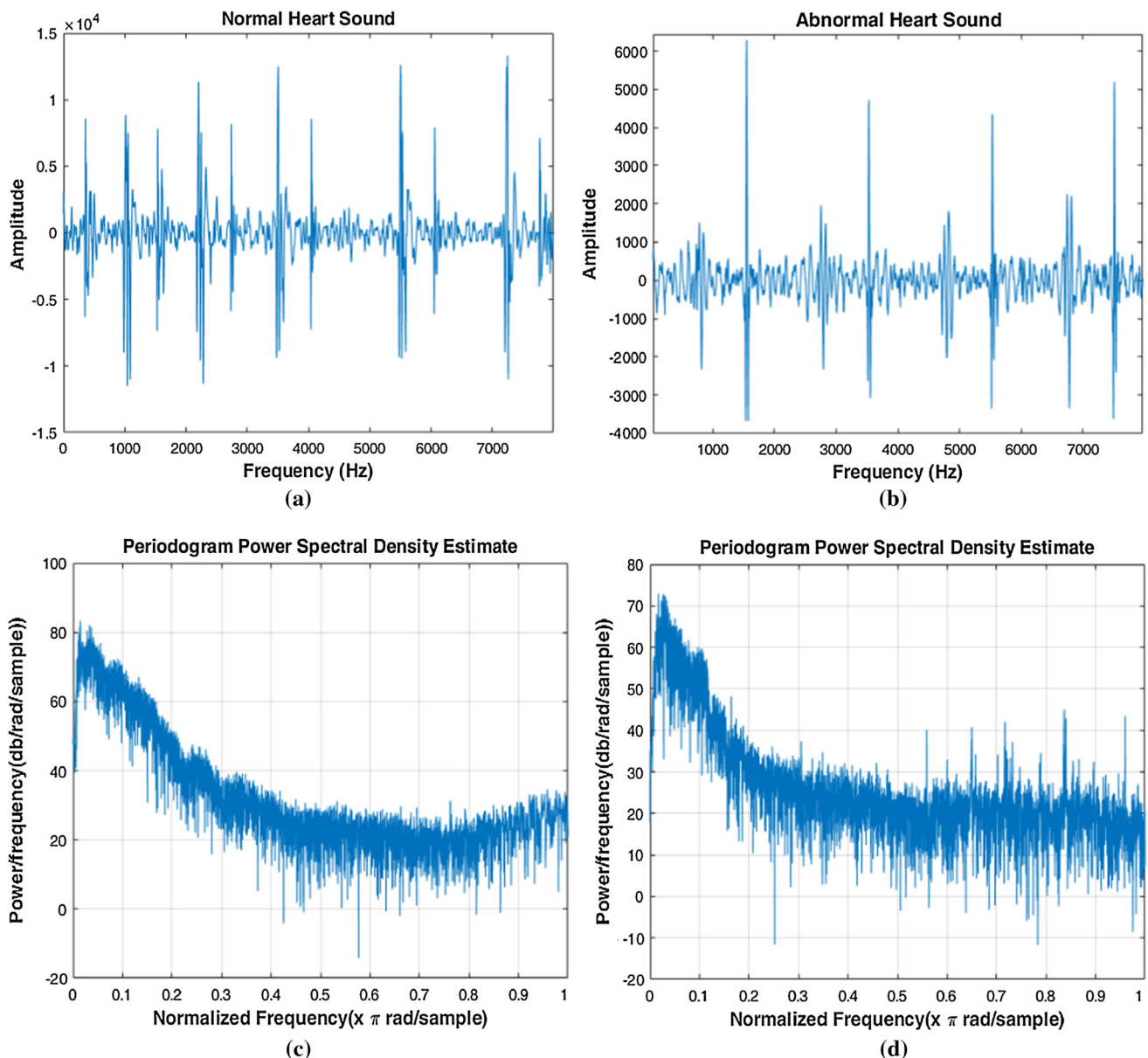

Fig. 1 a "Normal heart sound" recording b "Abnormal heart sound" recording c Power Spectrum of normal heart sound d Power spectrum of abnormal heart sound
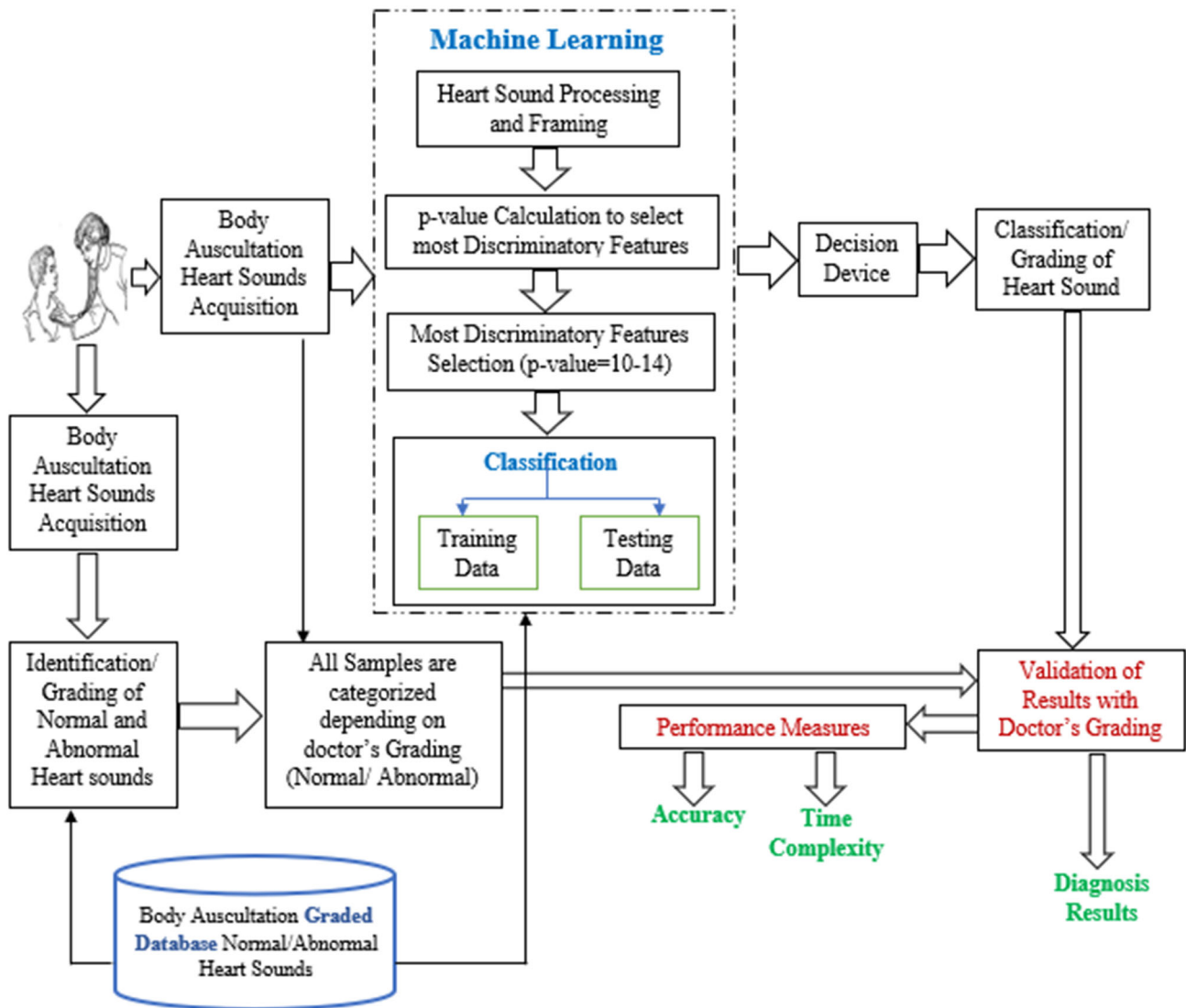
Fig. 2 Flowchart of the proposed method for diagnosis of cardiac diseases

- Processing and Framing, Extraction of Features
- Selection of Prominent Features and
- Supervised classification of normal/abnormal heart sounds.

3.1 Signal preprocessing

Automatic classification of heart sounds in normal and abnormal class is a challenging task due to the insignificant variance in heart sounds that are coming from normal or abnormal subjects. Keeping this in mind, there is a requirement to develop strategic methodology which can properly extract out the prominent and discriminatory features from the heart signals and which may further assist the machine learning-based classifiers in better classification of the heart sound in terms of normal and abnormal

heart sound signals. In the proposed work, audio samples of heartbeat are recorded with the help of PCG.

While recording, many noises and other artifacts are also recorded. So, denoising is the first and crucial step for proper analysis of the heart sound. For this purpose, bandpass filter is used. The range of the used bandpass filter is taken from 20 to 500 Hz. Lower cut-off frequency, i.e., 20 Hz is prescribed in the used database. The filtered signal retained all the information values corresponding to the above-mentioned frequency ranges. Other redundant and noisy information above 500 Hz are removed using the bandpass filter. Figure 3a presents the PCG recorded heartbeat signal. It contains different sort of noises which can corrupt the signal during analysis. The signal is passed through a band pass filter and the resultant signal is shown in Fig. 3b.

Fig. 3 a PCG recorded heart sound signal **b** Filtered heart sound signal
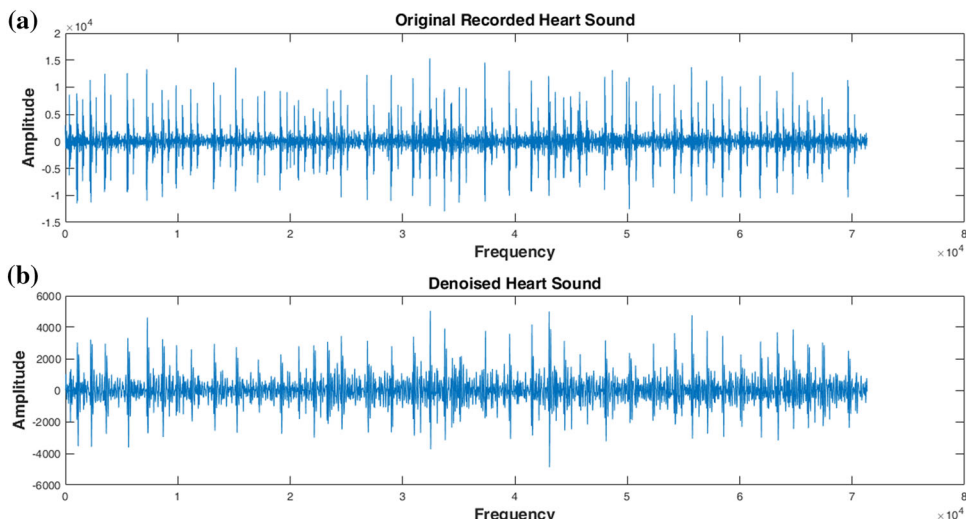
3.2 Features extraction

For proper classification of the heart sound signal, most prominent features are extracted from the recorded PCG heart sounds. In this developed algorithm, to increase the accuracy of the algorithm, each audio sample is split into non-overlapping frames and then various features are extracted from each frame of audio sample for improving the accuracy of the developed methodology.

Size of audio frame depends upon the size of audio sample in a proportionate optimized manner. All the heart sound signals are sampled at 2000 Hz. Bit rate of the audio samples is 16 bits per sample resulting into 32,000 bps. After extracting the features from each frame of the recorded signals, statistical features such as mean and standard deviation were calculated from all frames as a single representative of audio sample. Figure 4 shows flowchart of the framing of the filtered heart signal.

3.2.1 Extraction of the features from each frame of the heart sound

In this proposed work, six audio features (spectral centroid, zero crossing rate, energy entropy, spectral roll off, volume and spectral flux) are extracted from each frame of audio sample. Then, the statistical parameters are calculated as a single representative of that audio sample which eases the complexity of classification as well as reduces the computational time for classification.

(a) *Spectral Centroid* It is calculated in terms of weighted mean of frequency in the audio sample using Fourier transform.

$$\text{Centroid} = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{1}$$

where $x(n)$ defines weighted frequency of bin number. $f(n)$ is the center frequency of bin, n is number of bin whose range is 0 to $N-1$.

(b) *Energy Entropy* The heart sound is divided into K sub windows of specific duration. Energy entropy for each frame is calculated by the equation given below:

$$I_j = -\sum_{i=1\ldots k} \sigma_i^2 \log_2 \sigma_i^2 \tag{2}$$

where (σ_i^2) is the normalized energy calculated for each subwindow i.

(c) *Spectral Roll off* It is expressed as the Kth percentile of the total power spectral distribution in audio signal, where K is 85% or 95% spectral roll off of audio signal. The spectral roll off is also defined as the frequency F_t below which 85% is concentrated magnitude distribution. It can be calculated by the equation given below:

$$\sum_{n=1}^{R_t} N_t[n] = 0.85 \sum_{n=1}^{N} N_t[n] \tag{3}$$

where $N_t[n]$ is the amplitude of frequency component at frequency bin n and frame t. N is total number of frequency bin.

(d) *Spectral Flux* Spectral flux calculation is independent of total power and the phase consideration. It can be used for determining the timbre of audio. It can be calculated using the equation given below:

$$F_j = \sum_{k=0\ldots S-1} \left(N_{j,k} - N_{j-1,k} \right)^2 \tag{4}$$

where $N_{j,k}$: spectral energy of jth frame for Kth sample.

(e) *Zero Crossing Rate* It is defined as the rate of change of signs along the audio, i.e., it is the rate of change of signal

Fig. 4 Flowchart of the method of the framing of the filtered heart sound
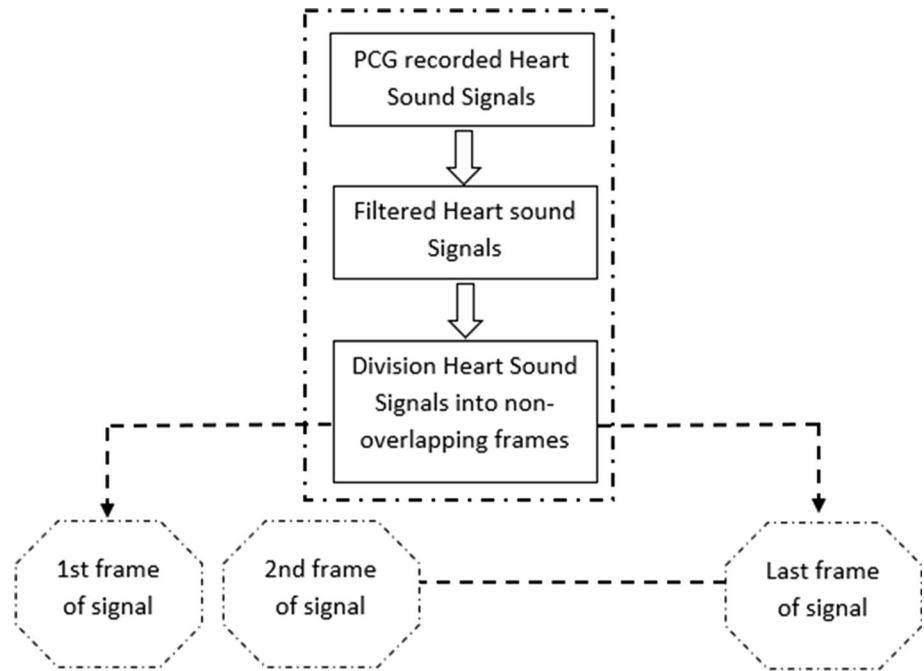
from positive value to negative value or vice versa. ZCR can be calculated as:

$$Z_j = \frac{1}{2S} \sum_{i=1\ldots S} |\mathrm{sgn}(x_i) - \mathrm{sgn}(x_{i-1})| \tag{5}$$

3.2.2 Statistical analyses of the extracted features

After extracting the different features frame-wise, statistical features like mean as well as standard deviation of all the extracted features are calculated. This is done so that the calculated statistical parameters act as a single representative of that audio sample which help in reducing the complexity of classification as well as the computational time at the time of classification step.

$$\text{Mean}\,(\mu) = \frac{\sum_r \sum_c \{J(a,b)\}}{A \times B} \tag{6}$$

$$\text{Standard Deviation}\,(\sigma) = \sqrt{\frac{\sum_a \sum_b \{J(a,b) - \mu\}^2}{A \times B}} \tag{7}$$

3.3 Discriminatory feature analysis and prominent feature selection

In the proposed work, machine learning algorithms are used for automated classification of normal/abnormal heart sounds for diagnosis of cardiac diseases. Accuracy and time complexity of classification algorithm depends upon the discriminatory features that are used in the classification algorithm. If all the extracted features are fed to the

classifier, then it may not only affect its accuracy but at the same time may also affect the complexity and CPU time for calculation. Thus, optimized feature selection is a crucial task as it enables the classifier to train faster on the correct subset of features. It also improves the accuracy of the developed model by making it easier to classify between normal and abnormal heart sounds.

Initially, twelve features are extracted from recorded heart sound sample which are further analyzed to find out most prominent features having discriminatory properties for accurate classification of heart sounds. In this proposed methodology, Wilcoxon rank sum test [22] is used which further checks discrimination for all the twelve extracted features. Wilcoxon rank sum test is a statistical hypothesis test which is used to find out the p-value. The most prominent properties are selected on the basis of p-value which is used for classification purpose. When the p-value is lower, then it shows higher discriminatory behavior among the features. Thus, in this paper, the p-values are considered as the standard to select discriminatory features for further classification. Based on the Wilcoxon rank sum rule, p-value is calculated for all 12 extracted features and most prominent features which have lower p-value are finally considered for classification.

Table 2 displays the calculated mean ± standard deviation along with p-value of each of the extracted 12 statistical features from the heart sound database. From Table 1 using p-value, it is quite visible that all features are not showing enough discrimination, which can used for further analysis. In order to reduce the time complexity of the classifiers and to obtain better accuracy, only most

Table 2 Extracted features from heart sounds and their corresponding p-value

S. no.	Extracted features	Normal samples (Mean ± SD)	Abnormal samples (Mean ± SD)	p-values	Selection of prominent feature
1	Mean: spectral centroid	0.0507 ± 0.0054	0.1188 ± 0.0327	3.27E − 14	√
2	Std. deviation: spectral centroid	0.0421 ± 0.0186	0.0201 ± 0.0059	1.52E − 10	×
3	Mean: energy entropy (Mean)	0.1306 ± 0.0061	0.0363 ± 0.0084	1.44E − 14	√
4	SD: energy entropy	0.0433 ± 0.0106	0.0742 ± 0.0214	6.00E − 09	×
5	Mean: spectral roll off	0.0413 ± 0.0028	0.0541 ± 0.0128	9.65E − 10	×
6	SD: spectral roll off	0.0285 ± 0.0103	0.0206 ± 0.0046	7.98E − 04	×
7	Mean: spectral flux	0.9595 ± 0.4901	2.5045 ± 0.5895	3.79E − 14	√
8	SD: spectral flux	1.8503 ± 0.6136	0.9526 ± 0.5482	9.73E − 11	×
9	Mean: zero crossing rate	0.0121 ± 0.0046	0.0528 ± 0.0160	1.43E − 14	√
10	SD: zero crossing rate	0.0231 ± 0.0126	0.0073 ± 0.0029	4.15E − 13	×

√ for selected prominent features and × for insignificant features for classification

Table 3 Selected discriminatory features and their corresponding p-value

Selected features	Normal samples (Mean ± SD)	Abnormal samples (Mean ± SD)	p-values
Mean: spectral centroid	0.0507 ± 0.0054	0.1188 ± 0.0327	3.27E − 14
Mean: energy entropy	0.1306 ± 0.0061	0.0363 ± 0.0084	1.44E − 14
Mean: spectral flux	0.9595 ± 0.4901	2.5045 ± 0.5895	3.79E − 14
Mean: zero crossing rate	0.0121 ± 0.0046	0.0528 ± 0.0160	1.43E − 14

prominent features are considered which have lower p-value. These discriminatory features are then fed to different classifiers which are used in the proposed algorithms.

Table 3 shows the selected 4 prominent features out of the 12 features extracted and their corresponding p-values subjected to classification. Another factor which supported Wilcoxon rank sum (p-value) method to find out discriminating behavior of feature vector is box plot of features. It gives the clear picture of discrimination of feature vector whether they can be used for classification purpose or not. Box plot helps in providing a better view, regarding the extent to which a feature is discriminative in nature.

Figure 5 presents the graphs of box plot of the selected prominent features which are further used for the purpose of classification. From the figure, it is clearly observed that the selected features show the enough discrimination which further helps in accurate categorization of abnormal and normal heart sounds using machine learning. Figure 5 shows the graphical distribution of selected features from normal and abnormal heart sounds. It can be seen in Fig. 6 that selected features are more significant to classify abnormal/normal heart sounds and there is a significant gap between the values of the features of two classes under consideration.

3.4 Train different machine learning classifier for the classification of recorded heart sounds

On the basis of Wilcoxon rank sum test, four most prominent features are fed to different classifiers for classifying recorded heart sounds as normal or abnormal class. In this proposed algorithm, different machine learning-based classifiers like Naïve Bayes, support vector machine, random forest and k-nearest neighbor are used. Among these classifiers, the best performing classifier is selected for classification.

Table 4 shows the different classifiers with different parameters used for classification. The main control parameters of the classification algorithms used for the proper classification from recorded heart sounds are explained below:

Kernel function that is used in SVM plays an important for proper training of the SVM. Then, using the grading of the recorded sounds provided by the doctors, support vectors are trained as normal and abnormal. The performance of the algorithm classification is for all possible types of kernel and finally the best parameter is selected for classification using test samples of recorded heart sounds. Figure 7 presents the step wise flowchart of SVM classifier
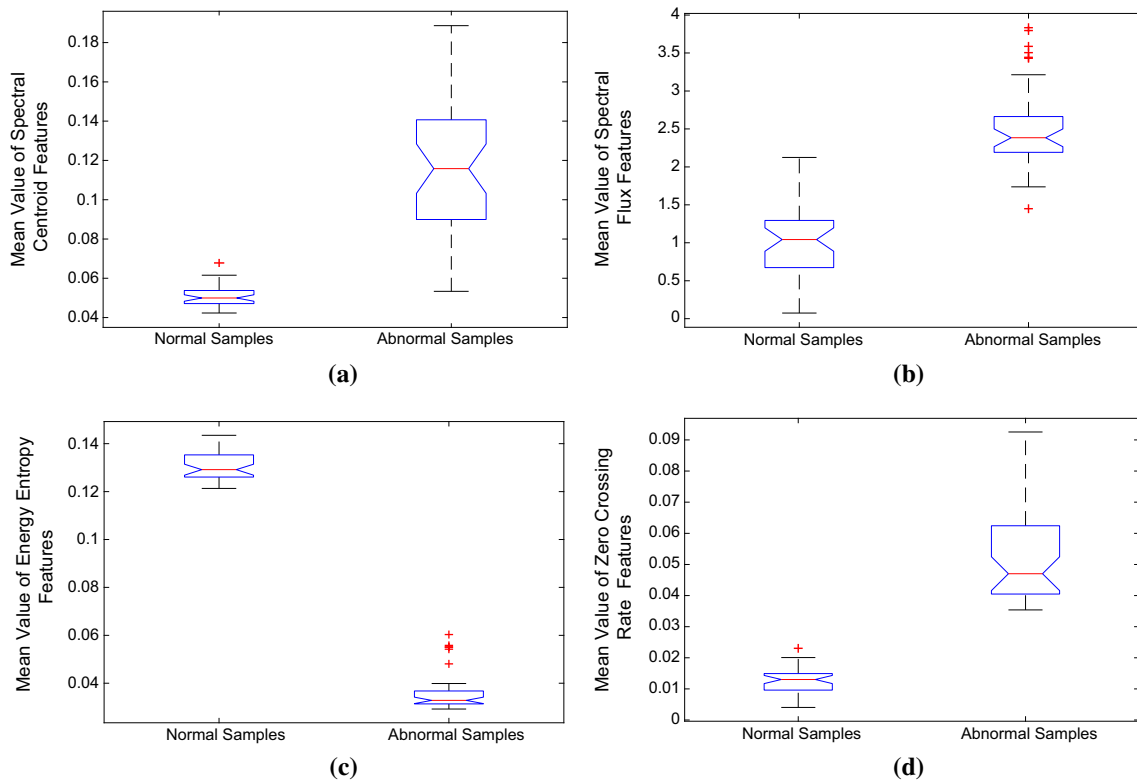
Fig. 5 Box plots showing discriminating behavior of selected features. **a** Mean: spectral centroid **b** Mean: spectral flux **c** Mean: energy entropy **d** Mean: zero crossing rate

which is used to classify healthy/unhealthy recorded heart sounds.

K-Nearest Neighbors algorithm classifies a data point on the basis of its neighbors. It is based on the feature similarity, and its distance measure metrics is selected according to the available training dataset. Choosing the right value of "*k*" in the processing is known as parameter tuning. It is very important for better accuracy obtained by *k*-NN algorithm. Figure 8 shows the flow diagram of the *k*-NN classification algorithm which is used in this work to classify between healthy and unhealthy heart sounds.

Naïve Bayes classifier works on Bayes rule and probability measurement. Classifier is trained using training dataset of heart sounds and best parameter setting is selected. On the basis of probability, best prediction is considered for classification of test sample of heart sound. Figure 9 presents flow diagram of Naïve Bayes classification algorithm used for machine learning.

In random forest classifier, the main control parameter is the number of decision trees. Using the training data with a set of features and targets, the algorithm of decision tree is generated with some specific set of rules. Performance of the random forest classifier is checked on the same set of rules and can be used to check the performance of the classifier on the testing data. Figure 10 shows the flow diagram of the random forest classifier used in this work.

To measure the performance of classifier, different parameters like sensitivity, specificity and accuracy [27] and Equal Error Rate [28] are calculated as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{9}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{10}$$

$$\text{Equal Error Rate (ERR)} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{11}$$

where TP is defined as true positive which presents total number of recorded heart samples from abnormal category which are classified as abnormal, TN is defined as True Negative which means the total number of recorded heart samples from normal category which are classified as normal, FP means False Positive which is used to expressed the total number normal heart samples which are classified as abnormal while FN means False negative which is used to defined the total number of abnormal heart samples which are classified as normal.

The performance of classifier is measured at various possible parameters, and best classification performance is considered for the classification of heart sounds.
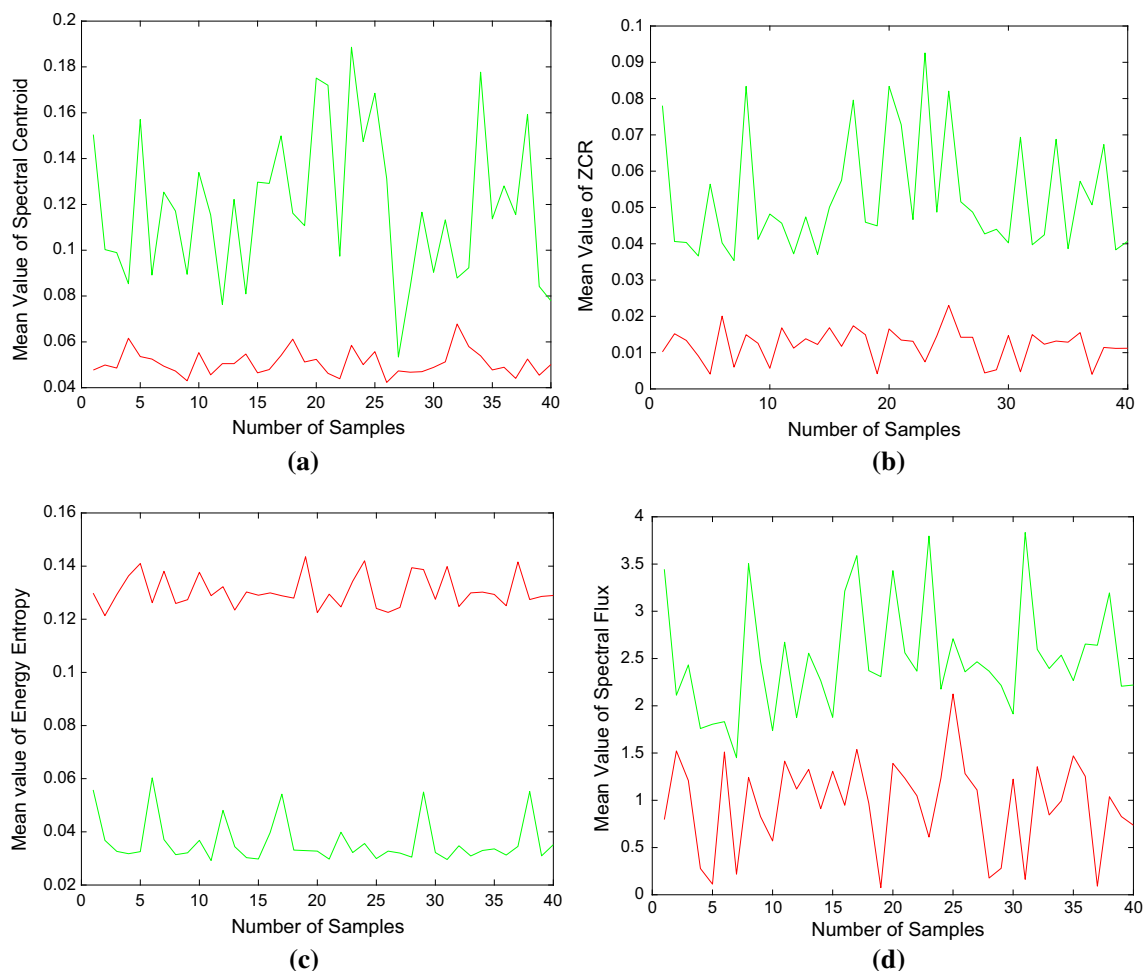
(a)

(b)

(c)

(d)

Fig. 6 Plots of selected prominent features from normal/abnormal heart sounds (Red Line Plot defines the extracted values of normal heart sounds. Green Line Plot defines extracted values of abnormal heart sounds) **a** Mean: Spectral Centroid **b** Mean: Spectral flux **c** Mean: Energy Entropy **d** Mean: Zero Crossing Rate

Table 4 Considered setting parameters for classifiers

S. no.	Classifiers	Setting parameters used for classification
1.	Support vector machine [23]	Linear kernel
2.	Naïve Bayes [24]	Normal distribution
3.	Random forest [25]	Decision trees = 75
4.	k- nearest neighbors [26]	Number of nearest neighbors = 5
		Distance = Cosine, Rule = Nearest

4 Experimental results and discussion

In the proposed methodology, the experiments are accomplished on a total of 500 heart sounds recorded from healthy patients as well as from those patients who are suffering from some sort of cardiac diseases. No overlapping audio sample is used in the experiment for training and testing purpose. Table 5 gives a detail information about the distribution of 250 Normal and 250 abnormal

heart sounds which are used to train the classifiers and for testing of the developed algorithms.

Table 6 presents the performance of different classifiers which are used to check the performance of the trained algorithm. Table shown is divided into two sections: first one describes the performance of the classifiers without feature selection, while the next shows the performance when the prominent feature are selected. From the table, it is clearly visible that all four classifiers achieved more than 90% accuracy for classification of PCG recorded heart

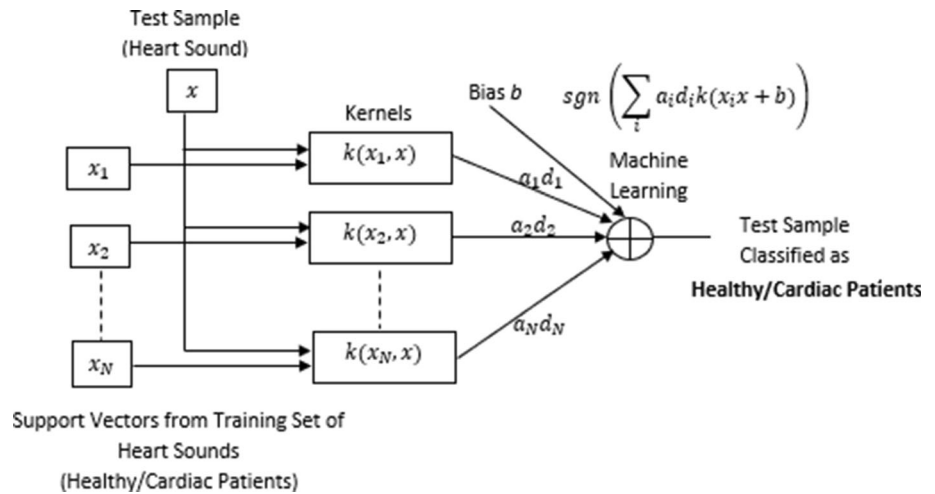**Fig. 7** SVM classification used to classify the recorded heart sounds as healthy or unhealthy heart sound

Test Sample (Heart Sound)

x

Kernels

Bias b $sgn\left(\sum_i a_i d_i k(x_i x + b)\right)$

x_1 $k(x_1, x)$ $a_1 d_1$ Machine Learning

x_2 $k(x_2, x)$ $a_2 d_2$ Test Sample Classified as **Healthy/Cardiac Patients**

x_N $k(x_N, x)$ $a_N d_N$

Support Vectors from Training Set of Heart Sounds (Healthy/Cardiac Patients)

Fig. 8 Flow of *k*-NN classifier used to classify the recorded heart sounds as healthy or unhealthy heart sound

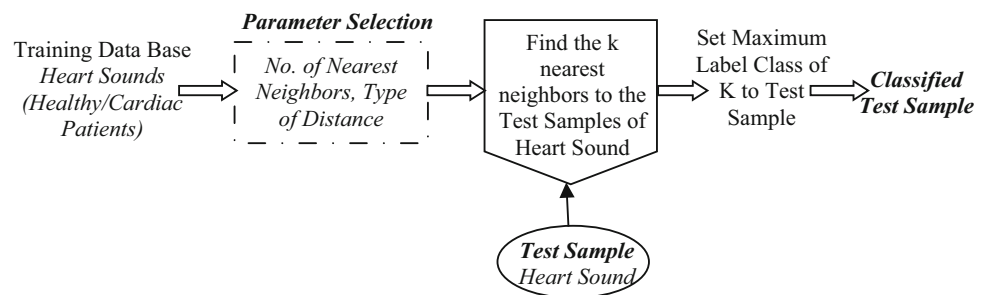*Parameter Selection*

Training Data Base *Heart Sounds (Healthy/Cardiac Patients)* → *No. of Nearest Neighbors, Type of Distance* → Find the k nearest neighbors to the Test Samples of Heart Sound → Set Maximum Label Class of K to Test Sample → *Classified Test Sample*

Test Sample Heart Sound

Fig. 9 Flow diagram of Naive Bayes classification algorithm

Training Set *Heart Sounds (Healthy/Cardiac Patients)*

Test Set *Heart Sounds*

→ *Bayes Rule (Normal Distribution)* → *Check the probability of test data to belonging in possible classes Healthy/Cardiac Patients)* → **Best Prediction for test Sample** *Healthy/Cardiac Patients)*
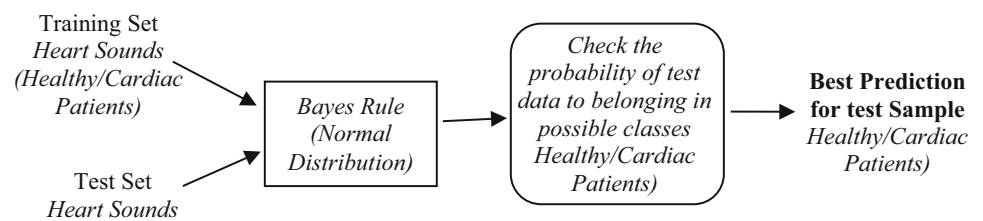
Fig. 10 Flow diagram of random forest classifier algorithm

Multiple Decision Tree is created on the basis of Random Vectors

Training Data (Heart Sounds: Healthy/Cardiac Patients)

Test Sample

Creation of Random Vectors → Randomize → DecisionTree-T$_1$ / DecisionTree-T$_2$ / DecisionTree-T$_n$ → *Results of Combine Decision Tree* → Average Value → *Test Sample Prediction*
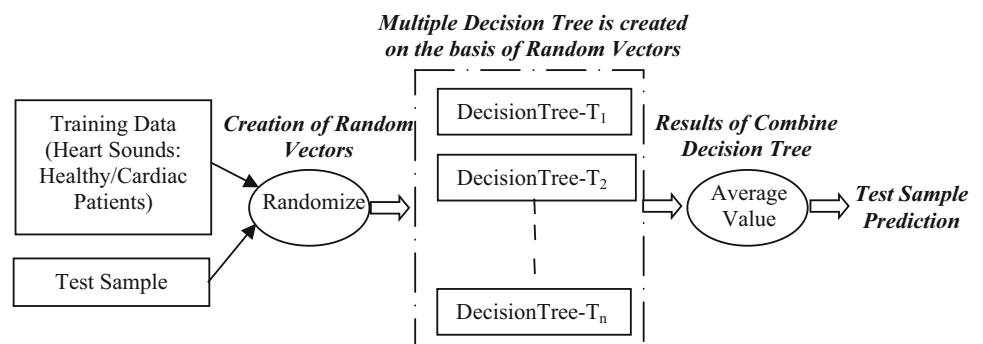
sounds as normal and abnormal. The best accuracy of 97.78% is provided by the SVM classifier with specificity of 97.78%, sensitivity of 98.46% and equal error rate (EER) of 2.22% for the diagnosis of cardiac disease which can be considered as potentially an important contribution for screening of cardiac diseases.

Tenfold cross-validation is performance to check the accuracy of each of the used classifier. In tenfold cross-validation, the dataset is nearly divided into k-equal segments which is called folds. Then ten iterations of training and testing are done in such a way that in each iteration one-fold is held for testing purpose while the remaining

Table 5 Heart sounds database used for experiments

Heart sounds	No. of heart sounds used for classifier training	No. of heart sounds used for testing purpose	Total samples
Normal	160	90	250
Abnormal	160	90	250
Total	320	180	500

Table 6 Comparative performance of different classifiers for normal/abnormal heart sounds with/without feature selection

Classifiers	TP	FP	TN	FN	Sensitivity (in %)	Specificity (in %)	Accuracy (in %)	EER (equal error rate)	CPU time (s)
Performance of classifiers without feature selection									
SVM	87	6	84	3	100%	93.33	95	5	0.9700
Naïve Byes	85	10	80	5	94.44%	88.88	91.66	8.34	0.9800
Random forest	85	7	83	5	94.44%	92.22	93.33	6.67	2.6400
k-NN	86	6	84	4	95.55%	93.33	94.44	5.56	2.3500
Performance of classifiers with prominent feature selection									
SVM	88	2	88	2	98.46%	97.78	97.78	2.22	0.8900
Naïve Byes	86	3	87	4	98.46%	95.55	96.67	3.33	0.8300
Random forest	86	2	88	4	92.30%	95.55	97.78	2.22	2.4500
k-NN	86	5	85	4	98.46	94.44	95	5	2.1800

ninefolds are used for training. In this algorithm, the data are arranged in such a way that in each fold classes of both heart sound samples are comprised equally. Table 7 presents tenfold cross-validation is performance to check the accuracy of each of the used classifier. Accuracy of each classifier is calculated by taking the average of the total accuracy of tenfolds. In tenfold cross-validation, the best classifier is SVM with the accuracy of 95.62%. On evaluating or comparing all performed classifiers, classifiers with prominent features are taken into consideration as they are providing best results.

Figure 11 presents the receiver operating characteristic (ROC) curve for all the used classifiers that are SVM, random forest, Naïve Bayes and k-NN, in a single plot. It can be seen in Fig. 10 that for all four classifiers, ROC curve is near to (0, 1) coordinate of the ROC space which is the indication of best possible prediction.

It is also clear in Fig. 11 that there is no significant difference in ROC curve for all used classifiers while classifying the normal and abnormal heart sounds. The area under the ROC curve or "AUC" (Area under curve) was calculated from ROC curve for all four classifiers. The value of AUC was 0.9827 for SVM classifier, 0.9956 for k-NN classifier, 0.9870 for Naïve Bayes classifier and 0.9912 for random forest classifier which shows the perfect classification. The EER is the point on the ROC curve where negative and positive rates are equal. While developing a system, this value is kept as small as possible.

Figure 12 is the graphical representation of comparative performance of classifiers in terms of accuracy with/without prominent feature selection and using tenfold cross-validation. Similarly, Fig. 13 presents the comparative performance of all the used classifiers in terms of time

Table 7 Comparative performance of tenfold cross-validation of different classifiers for normal/abnormal heart sounds

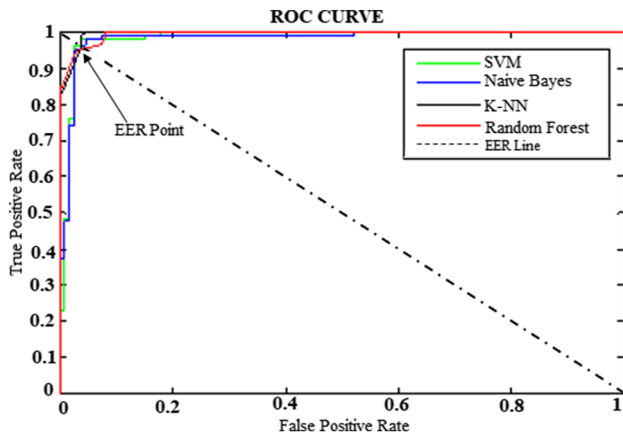| Classifiers	Sensitivity (in %)	Specificity (in %)	Accuracy (in %)	EER (equal error rate)	CPU time (in s)
SVM	95.73	94.78	95.62	4.38	0.9100
Naïve Bayes	91.44	96.34	94.00	6	0.9300
Random forest	96	93.48	95.17	4.83	2.2400
k-NN	96.51	91.45	94.00	6	2.0500

Fig. 11 ROC curve for random forest, Naïve Bayes, SVM and *k*-NN classifier

Fig. 13 Graphical representation of comparative execution time in terms of accuracy with/without feature selection

complexity with/without feature selection and using tenfold cross-validation.

The performance of different classifiers without feature selection, with prominent feature selection and with tenfold cross-validation is shown in Table 6. It clearly indicates that performance of machine learning algorithms for automated diagnosis of cardiac diseases has been improved with prominent feature selection method. The SVM, Naïve Bayes and *k-NN* classifiers show the best performance in terms of more accuracy and less time complexity with prominent feature selection method. All four classifiers with feature selection method achieved accuracy more than 95% for classification of abnormal and normal heart sound classification which can be considered as encouraging.

The classification algorithms are implemented on a workstation with computing tool MATLAB R2013a installed. The workstation has a configuration of Intel® Core™ i5-4210U CPU@ 1.7 GHz processor, 4 GB RAM, 64-bit operating system, × 64-based processor. The
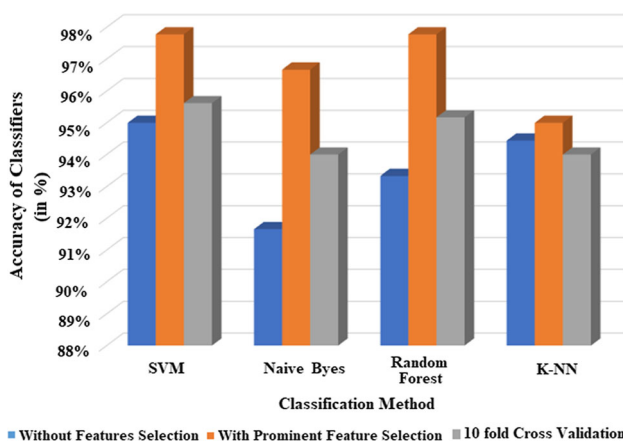
proposed automated system required less than 2.45 s time to test single sample which indicates that the proposed model can be considered as a significant contribution for diagnosis of cardiac diseases in real time.

5 Conclusion

In this developed algorithm, a supervised machine learning-based classification technique is applied to classify the PCG recorded heart sounds as normal/abnormal. Bandpass filter is used to remove the noise that gets recorded along with the heart sound. Presence of any sort of noise creates difficulty in extracting the useful information from the recorded signals of heart sound. Various statistical features are extracted from filtered heart sound signals. Wilcoxon rank sum is employed to find out the most prominent and discriminatory features among the extracted features. Through Wilcoxon rank sum, *p*-value of all the extracted features is calculated. In this paper, *p*-value is considered as an important parameter, which is used to filter out the discriminatory features for further analysis. This reduction in generalized features space to a more specific space will ensure better and more efficient classification. When the *p*-value is lower, it shows higher level of discriminatory behavior among the features. In the proposed work, the features were strategically considered which are likely to be affected during an abnormal heart functioning. Selection of most prominent features improved the performance of machine learning algorithms in terms of time complexity and accuracy. The proposed method achieved 97.78% accuracy with 97.78% sensitivity and 97.78% specificity for abnormal and normal heart sound classification when tested on heart sound database. The proposed model of automated diagnosis shows high accuracy with low time complexity as compared to existing method of

Fig. 12 Graphical representation of comparative classifier performance in terms of accuracy with/without feature selection and tenfold cross-validation
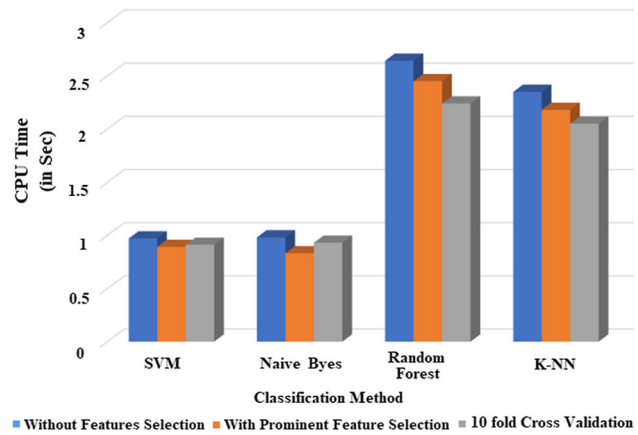
classification. Future work will be focused on the detection of various kind of the abnormal heart diseases. The algorithm can improve by detecting other body sounds such as sounds coming from lungs.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest

References

1. Mendis S, Puska P, Norrving B, World Health Organization (2011) Global atlas on cardiovascular disease prevention and control (PDF). World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization, pp 3–18. ISBN: 978-92-4-156437-3
2. Wilks S (1883) Evolution of the stethoscope. Popular Science 22(28):488–491
3. Leatham A (1975) Auscultation of the heart and phonocardiography. Churchill Livingstone, London
4. Azmy MM (2015) Classification of normal and abnormal heart sounds using new mother wavelet and support vector machines. In: 2015 4th international conference on electrical engineering (ICEE), Boumerdes, pp 1–3. https://doi.org/10.1109/intee.2015.7416684. December 2015
5. Mayorga P, Valdez J, Druzgalski C, Zeljkovic V (2016) Heart and lung sounds based events classification. In: 2016 global medical engineering physics exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE), Madrid, pp 1–6. https://doi.org/10.1109/gmepe-pahce.2016.7504614. April 2016
6. Maglogiannis I, Loukis E, Zafiropoulos E, Statis A (2009) Support vector machine-based identification of heart valve diseases using heart sounds. Comput Methods Prog Biomed 95:47–61
7. Gogineni J, Narayan JS, Rajeswara Rao D, Devi KP (2016) Development of efficient model for the assessment of heart risk stratification. Int J Pharm Bio Sci 7(3):1056–1060
8. Fu W, Yang X, Wang Y (2010) Heart sound diagnosis based on DTW and MFCC. In: 2010 3rd International congress on image and signal processing (CISP), Yantai, pp 2920–2923. https://doi.org/10.1109/cisp.2010.5646678
9. Gudadhe M, Wankhade K, Dongre S (2010) Decision support system for heart disease based on support vector machine and artificial neural network. In: 2010 international conference on computer and communication technology (ICCCT), Allahabad, Uttar Pradesh, pp 741–745. https://doi.org/10.1109/iccct.2010.5640377
10. Ghumbre S, Patil C, Ghatol A (2011) Heart disease diagnosis using support vector machine. In: International conference on computer science and information technology (ICCSIT'2011). Pattaya
11. Sheela CJ, Vanitha L (2014) Prediction of sudden cardiac death using support vector machine. In: 2014 international conference circuit, power and computing technologies (ICCPCT), Nagercoil, pp 377–381. https://doi.org/10.1109/iccpct.2014.7054771

12. Sonawane JS, Patil DR (2014) Prediction of heart disease using learning vector quantization algorithm. In: 2014 conference on IT in Business, Industry and Government (CSIBIG), Indore, pp 1–5. https://doi.org/10.1109/csibig.2014.7056973
13. Pedreira CE, Macrini L, Costa ES (2004) Input and data selection applied to heart disease diagnosis. In: Proceedings of international joint conference on neural networks, Canada, August 4, 2004
14. Vadicherla D, Sonawane S (2013) Decision support system for heart disease based on sequential minimal optimization in support vector machine. Int J Eng Sci Emerg Technol 4(2):19–26
15. Kumar D, Carvalho P, Antunes M, Paiva RP, Henriques J (2010) Heart murmur classification with feature selection. In: 2010 annual international conference of the IEEE engineering in medicine and biology, Buenos Aires, 2010, pp 4566–4569
16. Guraksin GE, Uguz H (2011) Classification of heart sounds based on the least squares support vector machine. Int J Innov Comput Inf Control 7(12):7131–7144
17. Hadi HM, Mashor MY, Mohamed MS, Tat KB (2008) Classification of heart sounds using wavelets and neural networks. In: 5th international conference on electrical engineering, computing science and automatic control, 2008. CCE 2008. Mexico City, 2008, pp 177–180. https://doi.org/10.1109/iceee.2008.4723403
18. Singh M, Cheema A (2013) Heart sounds classification using feature extraction of phonocardiography signal. Int J Comput Appl 77(4):13–17
19. Mandal S, Chatterjee J, Ray AK (2010) A new framework for wavelet based analysis of acoustical cardiac signals. In: 2010 IEEE EMBS conference on biomedical engineering and sciences (IECBES), KualaLumpur, pp 494–498. https://doi.org/10.1109/iecbes.2010.5742288
20. Hussain S et al (2013) Classification of heart sound based on multipoint auscultation system. In: 2013 8th international workshop on systems, signal processing and their applications (WoSSPA), Algiers, pp 174–179. https://doi.org/10.1109/wosspa.2013.6602357
21. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):215–220
22. Gibbons JD, Chakraborti S (2011) Nonparametric statistical inference. Springer, Berlin
23. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/bf00994018
24. Russell S; Norvig P (2003) [1995]. Artificial intelligence: a modern approach, 2nd ed. Prentice Hall, Upper Saddle River. ISBN: 978-0137903955
25. Ho TK (1995) Random decision forests (PDF). In: Proceedings of the 3rd international conference on document analysis and recognition, Montreal, QC, 14–16 August 1995, pp 278–282
26. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185. https://doi.org/10.1080/00031305.1992.10475879
27. Altman DG, Bland JM (1994) Diagnostic tests. 1: sensitivity and specificity. BMJ 308(6943):1552. https://doi.org/10.1136/bmj.308.6943.1552
28. Tharwat A (2018) Classification assessment methods. Appl Comput Inform. https://doi.org/10.1016/j.aci.2018.08.003

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.