Journal of King Saud University – Computer and Information Sciences xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Effect of Hilbert-Huang transform on classification of PCG signals using machine learning

Özkan Arslan [a,*], Mustafa Karhan [b]

[a] Department of Electronics and Communication Engineering, Tekirdağ Namık Kemal University, Çorlu 59860, Tekirdağ, Turkey
[b] Department of Electronics and Automation, Çankırı Karatekin University, Çankırı, Turkey

ARTICLE INFO

Article history:
Received 11 September 2021
Revised 29 November 2021
Accepted 27 December 2021
Available online xxxx

Keywords:
Phonocardiogram
Empirical mode decomposition
Time-frequency analysis
Genetic algorithm
Deep neural network

ABSTRACT

Heartbeat sounds are biological signals used in the early diagnosis of cardiovascular diseases. Digital heartbeat sound recordings, called phonocardiogram (PCG), are used in the determination and automatic classification of possible heart diseases. Healthy and pathological PCG signals are non-stationary signals and conventional feature extraction methods are insufficient in classifying these signals. In this study, PCG signals in healthy and four pathological categories are decomposed into intrinsic mode functions (IMFs) by Hilbert-Huang transform. Mel-frequency cepstral coefficient (MFCC) features were extracted from each mode to investigate the effect of the modes obtained by Hilbert-Huang transform on the classification of PCG signals. Genetic algorithm was used as feature selection method and k-nearest neighbor (KNN), multilayer perceptron (MLP), support vector machine (SVM) and deep neural network (DNN) machine learning methods were used as classifier. We have implemented multi classifications of five PCG classes (healthy, aortic stenosis, mitral stenosis, mitral regurgitation and mitral valve prolapse) by using 5-fold cross validation and 10×5-fold cross validation Data Analysis Protocol (DAP) framework. The results show that the DNN model provides the highest classification performance with 98.9% precision, 98.7% recall, 98.8% F1-score and 98.9% accuracy using 5-fold cross validation, and Matthews correlation coefficient of 0.981 using the DAP method.

1. Introduction

Heart sounds are biological signals that occur with the movement of heart valves and blood flow in heart. Heart sound signals are an important and effective biological signal used in the diagnosis of cardiovascular disorders (Maglogiannis et al., 2009; Ari et al., 2010). The measurement of heart sound signals on the skin in the area of the heart with a stethoscope is called a phonocardiogram (PCG). Heart sounds are non-stationary signals caused by the movement of the heart valves due to the blood flow entering and leaving the heart chambers. Heart valve disorders are a cardiovascular disorder, and failure to detect these disorders early can lead to blood coagulates inside the blood vessels, heart failure and fatal diseases. Therefore, the analysis, processing and classification of heart sound signals by applying signal processing techniques is an important and effective approach for detection of heart disorders (Zeng et al., 2021a; Zhong et al., 2020). The components of the heart sound signals are analyzed by signal processing methods and these components can be used to explain different heart diseases. Healthy and pathological heart sounds differ in terms of time and frequency components (Kumar and Saha, 2018). The movements of the heart valves produce sounds in the frequency range less than 2 kHz. The first part of PCG signal is produced by mitral and tricuspid valve and is denoted by S1. A complete cardiac cycle starting at S1 and ending at the beginning of the next S1 is the heart sound signal wave described as a heartbeat. The second component of PCG signal is denoted by S2 and is produced by the aortic and pulmonary valve. S1 usually has a longer duration and lower frequency than S2. The differences in the frequency of heart sound signals lead to murmurs which are often associated with heart valve defects. The murmurs are continuous vibrations caused by irregular blood flow in the cardiovascular system (Francisco, 1959).

* Corresponding author.
E-mail addresses: oarslan@nku.edu.tr (Ö. Arslan), mustafakarhan@karatekin.edu.tr (M. Karhan).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2021.12.019
1319-1578/

Please cite this article as: Ö. Arslan and M. Karhan, Effect of Hilbert-Huang transform on classification of PCG signals using machine learning, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2021.12.019

Classification of PCG signals into two categories as healthy and diseased has been the subject area of research and development for many years. With the existence of different types of PCG diseases taken by experts in clinical conditions, the need for automatic classification of disease types has arisen. There is a need for more effective signal analysis and classification methods since PCG signals with different disease types such as aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR) and mitral valve prolapse (MVP) are similar in nature. Therefore, this study focused on a signal processing approach that can distinguish the PCG signals belonging to different disease types which are in the non-stationary signal structure.

Classification of PCG signals as healthy and pathological has been studied for a long time (Chen et al., 2020; Li et al., 2020b; Potes et al., 2016). However, obtaining the correct diagnosis from heart sounds still emerges as an important problem. Numerous signal analysis and classification methods based on heart sound signals have been proposed with the presence of new signal processing techniques and artificial intelligence methods. Recently, many researchers have focused on feature extraction and classification studies on automatic heart diagnosis (Shuvo et al., 2021; Kobat and Dogan, 2021). Time-frequency methods are widely used due to the non-stationary nature of heart sounds. Bozkurt et al. (2018) proposed the use of mel-frequency cepstral coefficients (MFCCs) and spectrogram representation with convolutional neural networks (CNN) in detecting structural heart abnormalities from PCG signals. Thiyagaraja et al. (2018) classified sixteen type PCG signals with MFCCs and the Hidden Markov Model (HMM) with an average accuracy rate of 92.68%. Tschannen et al. (2016) proposed a wavelet-based deep CNN model for the healthy and pathological PCG signals for PhysioNet/CinC Challenge 2016 (Liu et al., 2016) data set, and the accuracy, sensitivity and specificity scores were obtained as 81.2%, 84.8% and 77.6%, respectively. Gonz et al. (2016) classified healthy and pathological heart sounds with 82.4% accuracy with temporal alignment techniques such as dynamic time warp (DTW) and spectral MFCCs of PCG signals. Noman et al. (2019) used the MFCC features obtained from heartbeat sounds in the PhysioNet/CinC 2016 data set with a deep CNN model and classified the PCG signals as normal and abnormal, and the accuracy of the model was determined as 89.22%.

Recently, wavelet transform (WT) and empirical mode decomposition (EMD), a time–frequency domain methods, are frequently used in the classification of heartbeat sounds (Meintjes et al., 2018; Ghosh et al., 2019; Kumar et al., 2020; Zeng et al., 2021b). Ghosh et al. (2019) used wavelet transform and random forest (RF) classifier to detect heart valve diseases from PCG signals, and the accuracy rates of their proposed method are 98.83%, 97.66%, 91.16% and 92.83% for healthy, aortic stenosis (AS), mitral stenosis (MS) and mitral regurgitation (MR), respectively. In another study based on the classification of heart valve diseases such as healthy, MR, MS and AS, local energy and entropy features were obtained by Chirplet transform and used with a multi-class composite classifier. It was determined that the proposed method has a sensitivity of 99.44%, 98.66% and 96.22% for AS, MS and MR categories, respectively (Kumar et al., 2020). Yaseen and Kwon (2018) classified cardiac audio signals using MFCCs and discrete wavelet transform (DWT) features with support vector machine (SVM), deep neural network (DNN) and k-nearest neighbor (KNN) classifiers for five PCG signal categories which include a healthy and four pathological classes. In case MFCC features are used with KNN, SVM and DNN classifiers, five classes of PCG signals were classified at 72.2%, 87.2% and 82.3% accuracy rates, respectively. Alkhodari and Fraiwan (2021) classified five types heart sound signal based on CNN and bi-directional long short-term memory (CNN-BiLSTM) model combined spatial and temporal features. They achieved an overall Cohen's kappa, accuracy, sensitivity, and specificity of 97.87%, 99.32%, 98.30%, and 99.50%, respectively. In a similar study, PCG signal and its first derivative are decomposed into a set of frequency sub-bands with a number of decomposition levels by using the tunable Q-factor wavelet transform method (Zeng et al., 2021). The experiments have been carried out on a publicly available PCG database, which include two types of classification, one for binary classification (normal vs. abnormal) and the other for multi classification (normal vs. aortic stenosis vs. mitral regurgitation vs. mitral stenosis vs. mitral valve prolapse). They reported the overall average accuracy for binary, four-class and five-class classification are reported to be 97.75%, 98.69% and 98.48%, respectively.

Many criteria and methods are used for the performance evaluation of classifier models. It is common to calculate accuracy and F1-score using the traditional cross-validation method. However, this method has limitations such as bias selection and overfitting effect. Therefore, Data Analysis Protocol (DAP), developed within the scope of the US-FDA led initiatives MicroArray and Sequencing Quality Control (MAQC/SEQC), is used as an alternative method for evaluating model performances (Maggio et al., 2018; Chierici et al., 2020). The DAP approach ensures selection bias and other over fitting effects, and guarantees honest performance estimates on external validation data subsets, and is performed as a 10x5 fold cross-validation scheme with the feature selection and ranking procedure (Fioravanti et al., 2018). The top-ranked features are selected recursively as kbest in each round. Model performances are calculated for the increasing number of best ranking features by the Matthews Correlation Coefficient (MCC), which provides the best obtain ability of a classifier's confusion matrix even in the multi-class case. Repeated experiments with random features and labels are performed with the best results.

In this study, we have proposed an automatic PCG classifications using several machine learning models with features that obtain by Hilbert-Huang transform. For this purpose, we have used KNN, MLP, SVM and DNN models to obtain higher prediction accuracies for PCG dataset including signals of normal (healthy), aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR) and mitral valve prolapse (MVP) patients. It is aimed to increase the efficiency of the conventional MFCC method, which is frequently used in feature extraction for non-stationary signals, by applying the Hilbert-Huang transform. The conventional MFCC features were extracted from PCG signals, which were decomposed into modes by Hilbert-Huang transform, and the effects of modes on classification performance were investigated. Genetic algorithm was used to select the most distinctive features for KNN, MLP and SVM methods. In the training and testing of the DNN model, the features obtained directly from the modes features were used. Also, all classifier models based on 10 repetitions of 5-fold cross validation become the core of an experimental setup designed according to the DAP framework. The training set goes through 10 rounds of 5-fold cross validation with Anova-F value as classification score and k-best as feature selection algorithm. In each round, several models are built using the performance measure MCC to increase the number of features ranked. The novelty and originality of proposed study are summarized as follows:

- The proposed model classifies five different PCG signals, one healthy and four pathological, instead of the binary classification frequently encountered in the literature.
- A new and effective stable feature generator is presented by Hilbert-Huang transform.
- It has been studied and compared with four different classifier models, namely KNN, MLP, SVM and DNN.
- A high accurate PCG sound classification model is presented and this model attained accuracy values of 98.9% by employing DNN classifier with 5-fold cross validation.

- The 10 × 5-fold cross validation DAP results show that the DNN classifier, with an MCC value of 0.981, is more successful than other classifiers on categorizing PCG signals.
- A high-accuracy decision support system has been proposed to experts for the automatic diagnosis and detection of patients with suspected cardiovascular disorders and follow-up.

The remainder of this work is organized as follows. In section 2, the proposed method for the classification of PCG signals are explained. In section 3, the experimental results and detailed explanations of the proposed approach are presented. In section 4 and 5, the discussions and conclusions of this paper are provided, respectively.

2. Materials and methods

2.1. Dataset

The PCG database used in the study consists of five categories which are healthy and four pathological sounds (Yaseen and Kwon, 2018). This PCG sound dataset is an open-access and is published on GitHub publicly. This dataset includes aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR), mitral valve prolapse (MVP) and normal (healthy). The number of PCG records in each category is 200, and the total number of PCG signals is 1000. The length of each category is 548.36, 451.31, 471.33, 496.22 and 476.27 s. The sampling frequency of PCG signals is 8 kHz. The healthy and pathological PCG signals collected in this database are filtered and standardized for analysis and processing.

2.2. Method

The stages of the proposed method are illustrated in Fig. 1. In the signal decomposition stage, the PCG signals, which include normal and four pathological signals, are decomposed into modes by EMD method and the first six modes were selected for each signal. In the feature extraction stage, MFCC features are obtained for each selected modes. Genetic algorithm was used to select the most distinctive features for KNN, MLP and SVM methods. And, the selected mode features are used in the training and testing of the deep neural network and the classification of PCG signals is provided.

2.2.1. Hilbert-Huang transform

Hilbert-Huang transform (HHT) is an effective signal processing method used in the analysis of non-stationary and non-linear signals. HHT is a combination of empirical mode decomposition (EMD) and Hilbert transform approaches. In the first step of HHT, the signals are decomposed into modes by EMD.

EMD is an adaptive method that is used extensively to decompose non-stationary signals (Arslan and Engin, 2019). The EMD separates the signals into modes from high to low frequencies. In the decomposition process with EMD, the signal itself is used as a basis function unlike Fourier transform and wavelet transform (Khaldi, 2012). Therefore, the EMD is an adaptive method compared to other decomposition methods. The coefficients in the modes obtained by the decomposition of the signal are called intrinsic mode functions (IMFs). Each of the IMFs in the modes represents different frequency components of the signal. Thus, the number of samples in each mode is fixed unlike the down-sampling based wavelet transform method. The flowchart of the EMD is illustrated in Fig. 2.

In Fig. 2, all extrema of original signal is defined in separating the signal into IMFs with the sifting process. Then, the envelopes of the extrema are calculated by cubic spline interpolation. In the next step, the average of the envelopes is defined and $h(t)$ is obtained and assessed as it ensures stopping criterion SD < 0.3 or two IMF conditions. Two conditions of the IMF are determined as the difference between the number of extrema and zero crossing is less than one and the average of envelopes is zero. The stopping criterion (SD) is expressed as in Eq. (1):

$$SD(i) = \sum_{t=0}^{N} \frac{|h_{i-1}(t) - h_i(t)|}{h_{i-1}^2(t)} \tag{1}$$

where $h(t)$ and i represent the IMF and mode number, respectively. The $h(t)$ that satisfies the IMF condition or the stopping criterion is considered to be the first $IMF(t)$.

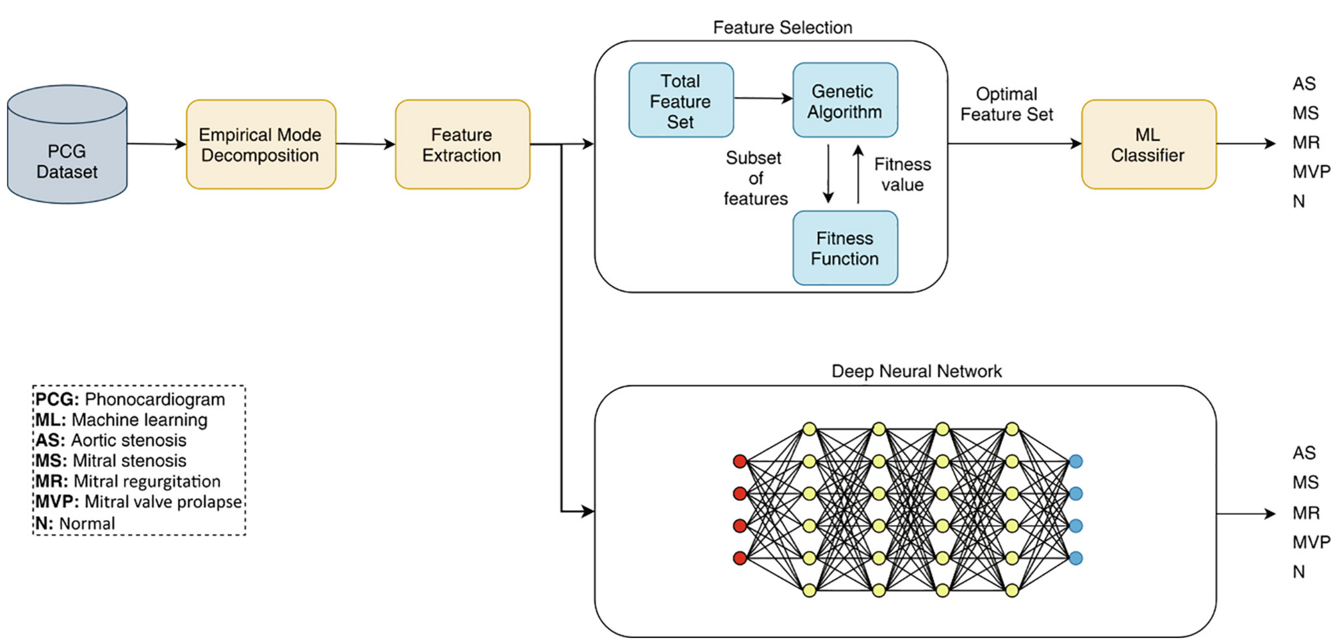

Fig. 1. Block diagram of the presented PCG sound classification models.

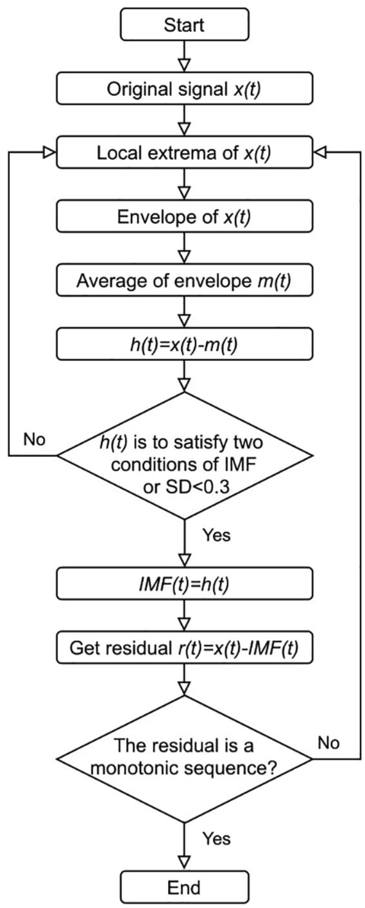

Fig. 2. The flowchart of EMD algorithm.

2.2.2. Feature extraction

Mel-frequency cepstral coefficients (MFCCs) are a method used extensively in sound processing and classification (Zheng et al., 2001; Sahidullah and Saha, 2012). The first step in extracting the MFCC features is to increase the amount of energy at high frequencies. The technique of MFCC basically involves discrete fourier transform (DFT), logarithm, mel-scale and discrete cosine transform (DCT) operations.

The sound signal is a non-stationary signal that changes over time and the sound must be examined within short segments. Short-term spectral measurements are typically performed at windows and it is ensured that the temporal properties of sounds are monitored. Hanning or Hamming window functions are generally used for windowing signal. The MFCC is expressed as in Eq. (2):

$$c(n) = \sum_{m=0}^{M-1} \log_{10}(s(m)) \cos\left(\frac{\pi n(m-0.5)}{M}\right); n = 0, 1, \cdots, C-1 \qquad (2)$$

where $s(m)$, $c(n)$ and M represent mel spectrum, cepstral and total mel filter number, respectively. And, C is the number of MFCC coefficients. The cepstrum plays an important role in improving performance in sound/audio recognition systems. Generally, the first 13 cepstral values are used in the sound/voice recognition system (Kim, 2013).

2.2.3. Feature selection

Feature selection is a very important technique in machine learning and requires heuristic processes to find the optimal machine learning subset. One of the most advanced algorithms for feature selection is the Genetic algorithm (Fröhlich et al., 2003; Tan et al., 2008; Babatunde et al., 2014). Genetic algorithm (GA) is an optimization method based on the natural selection process and work on a population of individuals to generate better approximations. The GA includes three basic operations: selection, crossover, and mutation. The selection describes which solutions are retained for further replication, while the crossover describes how new solutions are created from existing solutions. Mutation aims to bring diversity and innovation to the solution pool by randomly changing or turning off the solution bits. The basic structure of GA is shown in Fig. 3.

In the presented study, binary GA is applied to discard irrelevant features to find the best possible combination of features and create an effective model to increase the accuracy in classification of PCG signals. The initial population was randomly generated and updated at every iteration. Chromosome length is 13 and consists of 0′s and 1′s indicating absence and presence of features, respectively. In the machine learning model, the inverse of the squared error is accepted as the fitness function.

2.3. Machine learning methods

2.3.1. K-nearest neighbor

The basic principle of nearest neighbor methods is to find the nearest predetermined number of training samples to the new point and estimate the label. The number of samples can be a user-defined constant (k-nearest neighbor learning) or it can vary depending on the local density of points (Narendra, 1975). Distance can generally be any metric measure: for example, the standard Euclidean distance is the most common choice. Neighbor-based methods are known as non-generalized machine learning methods and classification is sample-based learning. The classification is calculated by a simple majority vote of each point's nearest neighbors, and a query point is assigned the data class with the most representative among the point's nearest neighbors (Laaksonen and Oja, 1996; Kramer, 2013).

2.3.2. Multilayer perceptron

Multilayer perceptron (MLP) is one of the methods used in the early stages of deep learning. MLP is a nonlinear function estimator for classification given a set of features $\mathbf{X} = x_1, x_2, \cdots, x_m$ and a target $\mathbf{y}$ by training on a dataset. MLP neural networks consist of an input layer to which input parameters are applied, an output layer

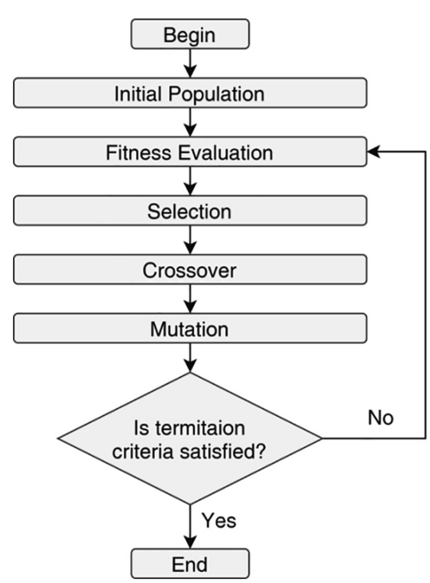

Fig. 3. Basic structure of Genetic algorithm.

that makes a prediction about the input, and an arbitrary number of hidden layers that are a computational tool between the input and output layer (Taud and Mas, 2018). MLP is generally used in supervised learning applications and it is provided to learn a model that gives the correlation between input and output. In the training phase, the parameters and weight coefficients of the model that minimize the error are obtained (Saravanan and Sujatha, 2018).

2.3.3. Support vector machine

Support vector machine (SVM) is basically a two-class classifier that fits a discriminating hyperplane between two classes (Hearst et al., 1998; Brereton and Lloyd, 2010). The optimal hyperplane is selected according to the maximum margin criterion and it is chosen to maximize the Euclidean distance to the nearest data points on each side of the plane. The data points closest to each separating hyperplane are known as support vectors (Mavroforakis and Theodoridis, 2006). The non-linear SVM can be expressed as (Bishop, 2006):

$$f(x) = \sum_{i=1}^{N} \alpha_i t_i K(x, x_i) + d \tag{3}$$

where $t_i\{+1-1\}$ are ideal output values. The support vectors x_i and their corresponding weights α_i and the bias term d are determined from a training set using an optimization process. Kernel function $K(x, x_i)$ can be expressed as:

$$K(x, x^T) = \phi(x)^T \phi(x) \tag{4}$$

where $\phi(x)$ is a mapping from the input space to the high-dimensional kernel feature space. The kernel function calculates the inner product of two vectors in the kernel feature space.

2.3.4. Deep neural network

Deep neural networks (DNN) are an effective method that has been used extensively in biomedical signal processing in recent years (Cao et al., 2018; Li et al., 2020b; Deperlioglu et al., 2020). Deep learning algorithms have very high computational power to process large numbers of data. DNN are a self-learning structure using distinctive features according to input. In this learning structure, the data itself can be used as distinctive inputs or can be used by obtaining important and effective features from input data. Deep learning structure includes input, output and hidden layers.

In the deep learning model, many layers consisting of neurons such as dense layer, dropout layer and soft-max layer are used. The rectified linear unit (ReLU) activation function in dense layers and softmax activation function in output layer are used extensively (Baydoun et al., 2020; Deng et al., 2020). The ReLU activation function used in the dense layer is mathematically expressed as:

$$h(a_j) = \max\{0, a_j\} \tag{5}$$

where a_j and h represent first hidden layer output and ReLU activation function, respectively.

The softmax is an activation function that allows the output to be represented in categorical ways and handles multiple classification problems. The softmax function is defined as follows:

$$\sigma(a_k) = \frac{e^{a_k}}{\sum_{j=1}^{K} e^{a_j}} \tag{6}$$

where a_k is calculated using the weight parameters. σ and K denote the softmax activation function and neuron number of the output layer, respectively.

2.4. Experimental setup

MATLAB2019b was used for feature extraction by Hilbert-Huang transform from PCG sound signals and Python programming language was used for training and testing machine learning methods. The neighbor value k for the KNN model was set to 5. SVM model was trained with penalty term (C = 1), gamma value (0.001) and 3rd degree poly kernel function for all experiments. MLP model was trained with random initialization weights by adaptive moment estimation (ADAM) optimizer and ReLU activation function was used. In MLP models, hidden layer size, learning rate and number of epochs were set to 100, 0.001 and 300, respectively. In all experiments, the parameter setting of DNN model is shown in Table 1.

All datasets were randomly split into two independent datasets with 80% and 20% for training and testing, respectively. In the evaluation of the models, the k-fold cross validation method was preferred. Results were obtained according to the k-fold value (k = 1–5) shown in the Fig. 4, and these results were averaged.

To achieve high predictive power and limit the effect of overfitting, the experimental setup was performed Data Analysis Protocol (DAP) developed within the US-FDA led initiatives MAQC/SEQC, which led to the development of predictive models for efficient data analysis. All classifier models based on 10 repetitions of 5-fold cross validation become the core of an experimental setup designed according to the DAP framework (Fioravanti et al., 2018) shown in Fig. 5. In the DAP method, the dataset is first split into a non-overlapping training set and test set, and in subsequent experiments, the training set size is 80% of the original dataset. Then, the training set goes through 10 rounds of 5-fold cross-validation with Anova-F value as classification score and k-best as feature selection algorithm. In each round, several models are built using the performance measure MCC (Chicco et al., 2021) to increase the number of features ranked. The number of ranked features for each model is increased to 25%, 50%, 75% and 100% of the total features. MCC is an elective choice to efficiently combine the confusion matrix of a classification task and evaluate the results of classifiers even when the classes are unbalanced (Chicco and Jurman, 2020). MCC values range from −1 to 1; where 1 indicates perfect classification, −1 indicates perfect misclassification. The lists of ranked features produced within the cross-validation scheme are then combined into a single ranked list. A portion of the combined list of ranked features corresponding to the higher MCC value is selected as the most appropriate set of distinguishing features for classification task. Finally, the same methodology is applied several times to samples of the original dataset after selecting random features instead of selecting random labels on basis of model performances.

2.5. Performance metrics

The precision, recall, F1-score, accuracy and Matthews correlation coefficient (MCC) criteria are used extensively in testing the classifier performance. These measures are determined from the

Table 1
Parameter setting of DNN model.

DNN Parameter	Value
Learning rate	1e-5
Number of epochs	100
Batch size	32
Dropout rate	0.35
Hidden layers	64–64-16
Hidden layer activation	ReLU
Optimizer	ADAM
Loss	categorical cross-entropy
Output activation	softmax

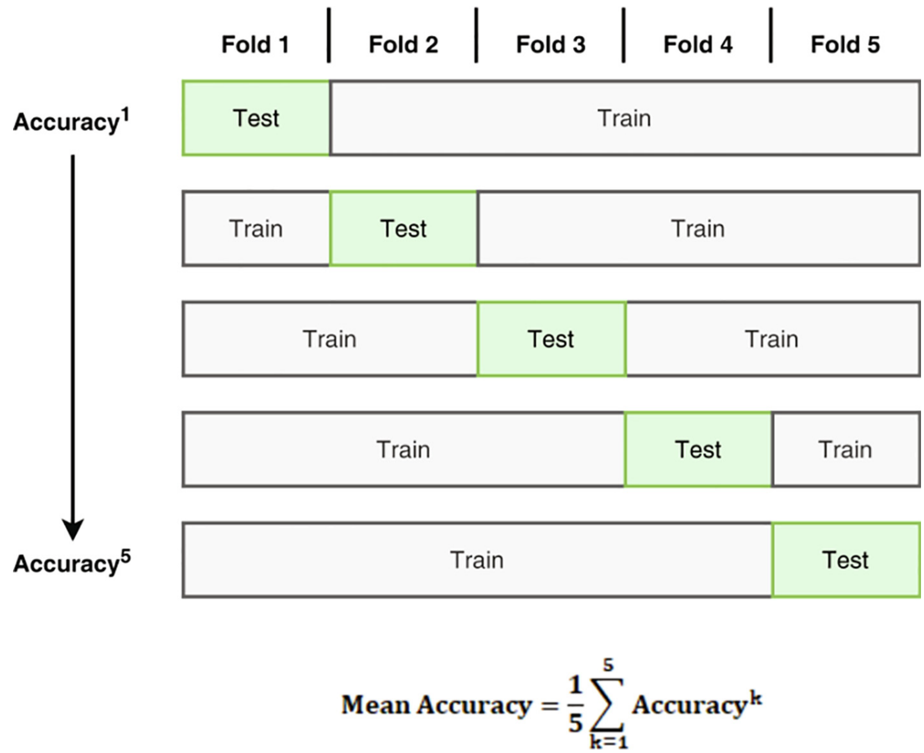

Fig. 4. Display of testing and training datasets for five-fold cross validation.

classifier's true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values. The confusion matrix used to determine the classifier performance is shown in Fig. 6.

The precision, recall, F1-score and accuracy are given in Eqs. (7)-(10):

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

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

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP \times FP)(TP \times FN)(TN \times FP)(TN \times FN)}} \tag{11}$$

3. Results

In this study, we perform Hilbert-Huang transform (HHT) based mel-frequency cepstral approach with four machine learning models for PCG sound classification. The pathological (AS, MR, MS and MVP) and normal (healthy) PCG signals were first decomposed into six modes by the EMD method which is the first step of HHT. The

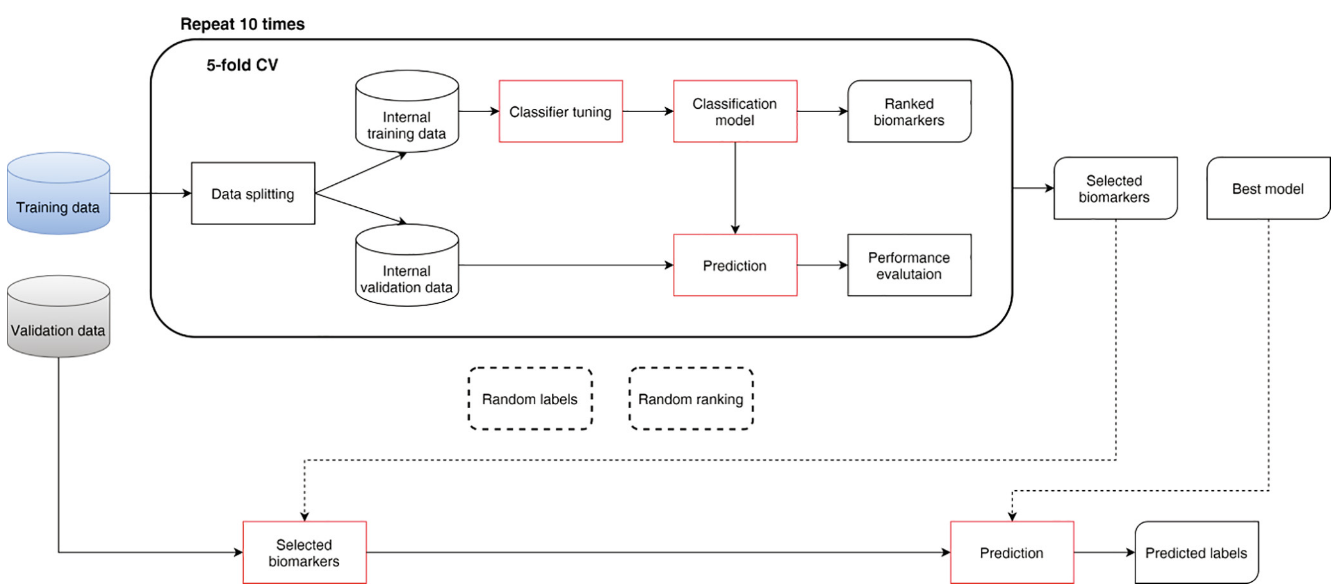

Fig. 5. Data Analysis Protocol (DAP) framework for the experiments.

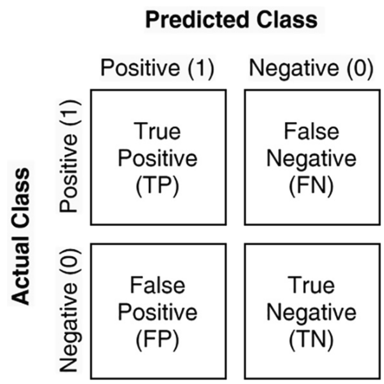

Fig. 6. Confusion matrix.

time waveform of PCG sounds, Hilbert spectrum of the signal using the calculated IMFs and visualize the instantaneous energy distributions of the signals are shown in Fig. 7.

As can be seen from Fig. 8, all PCG signals are limited in the frequency range from 0 Hz to 500 Hz. The energy distribution of different disease types and healthy PCG sounds varies according to time and frequency. Considering the time–frequency properties of PCG sounds according to the modes, MFCC approach was used as the feature extraction method. A total of 78 (6x13) cepstral coefficients, 13 coefficients for each mode, were used in the training and testing of the models. The boxplots for 5th MFCC features for mode 1–6 are shown in Fig. 8 (a)–(f). It is clear that MFCC features have lower mean value in mode 1 and 6 for MR, in mode 2 and 5 for normal (healthy), in mode 3 for MVP, and in mode 4 for MS classes.

Five-fold cross-validation has been used in this study performed with four different models that are KNN, MLP, SVM, and DNN. In calculating the performance of the models, 80% of the data is reserved for training and the remaining 20% for testing. GA was used to select the most distinctive features for KNN, MLP and SVM models. GA with an initial population of 1000*13, with a mutation rate of 0.25%, and the single point crossover is implemented. The results of the GA algorithm further aid us to conclude that some of the features in modes are strongly associated with the PCG classes. For the dataset of size 1000*13, the genetic algorithm is applied, optimized feature subset is obtained, and results are compared. The accuracy results of the each PCG category based on four machine learning methods for each mode (from 1 to 6) is shown in Fig. 9. As can be seen from Fig. 9, the classification performance of the models decreased as the number of modes increased. These results reveal that in high modes the signals are located in the low frequency bands and the MFCC coefficients are in the same frequency band. It can be said that the mode 1 and 2 features provide higher accuracy values among values of other modes for all PCG

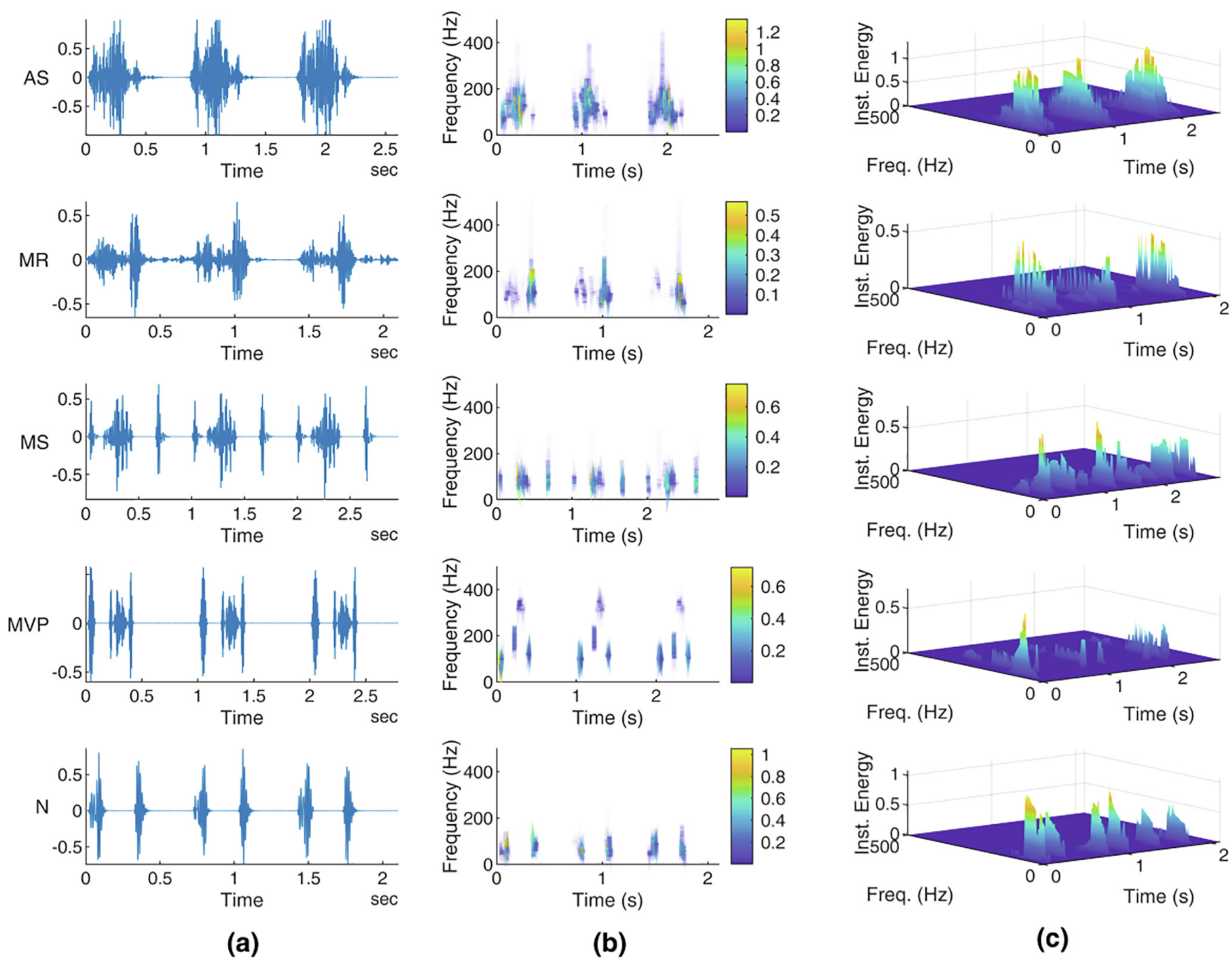

Fig. 7. PCG signals (a) time waveform (b) Hilbert spectrum, (c) instantaneous energy distributions of IMFs.

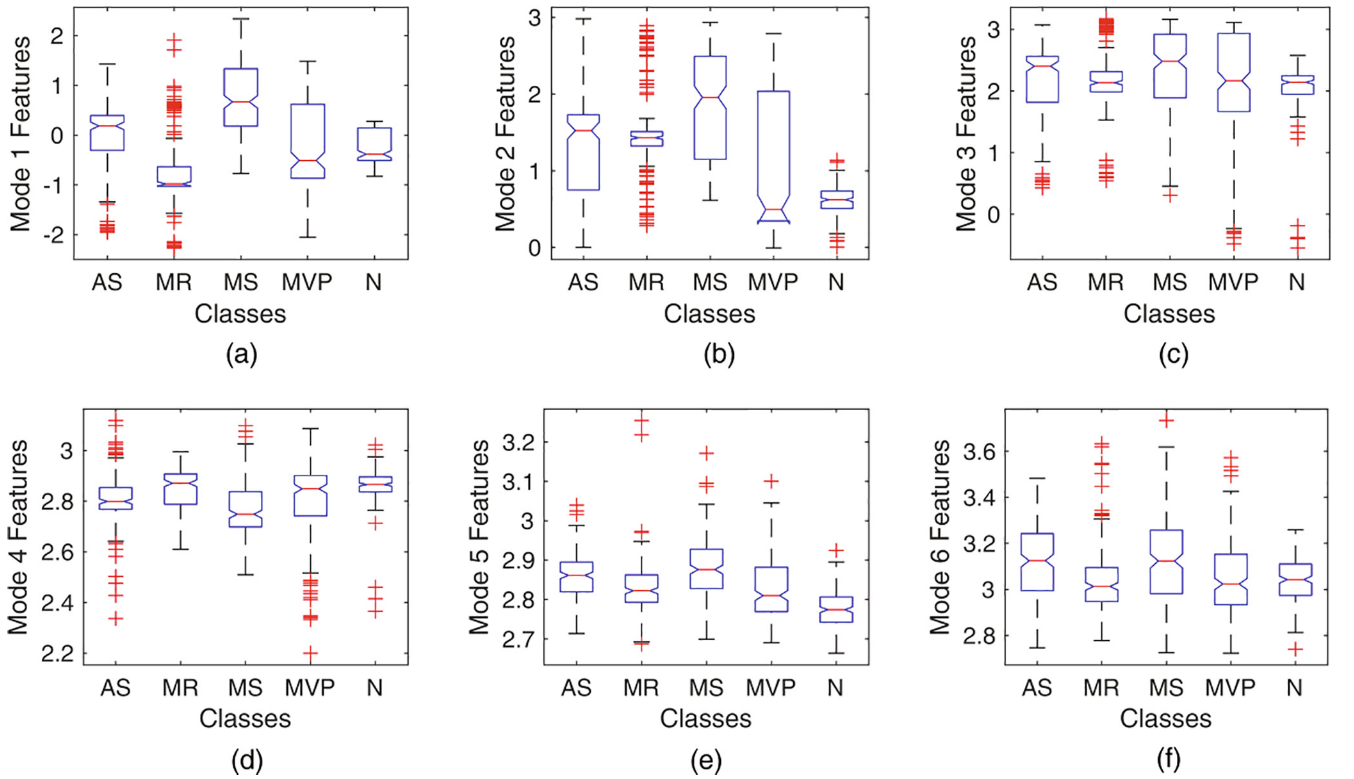

Fig. 8. Boxplot of features for all PCG classes (a) 5th MFCC feature for mode 1 (b) 5th MFCC feature for mode 2 (c) 5th MFCC feature for mode 3 (d) 5th MFCC feature for mode 4 (e) 5th MFCC feature for mode 5 (f) 5th MFCC feature for mode 6.

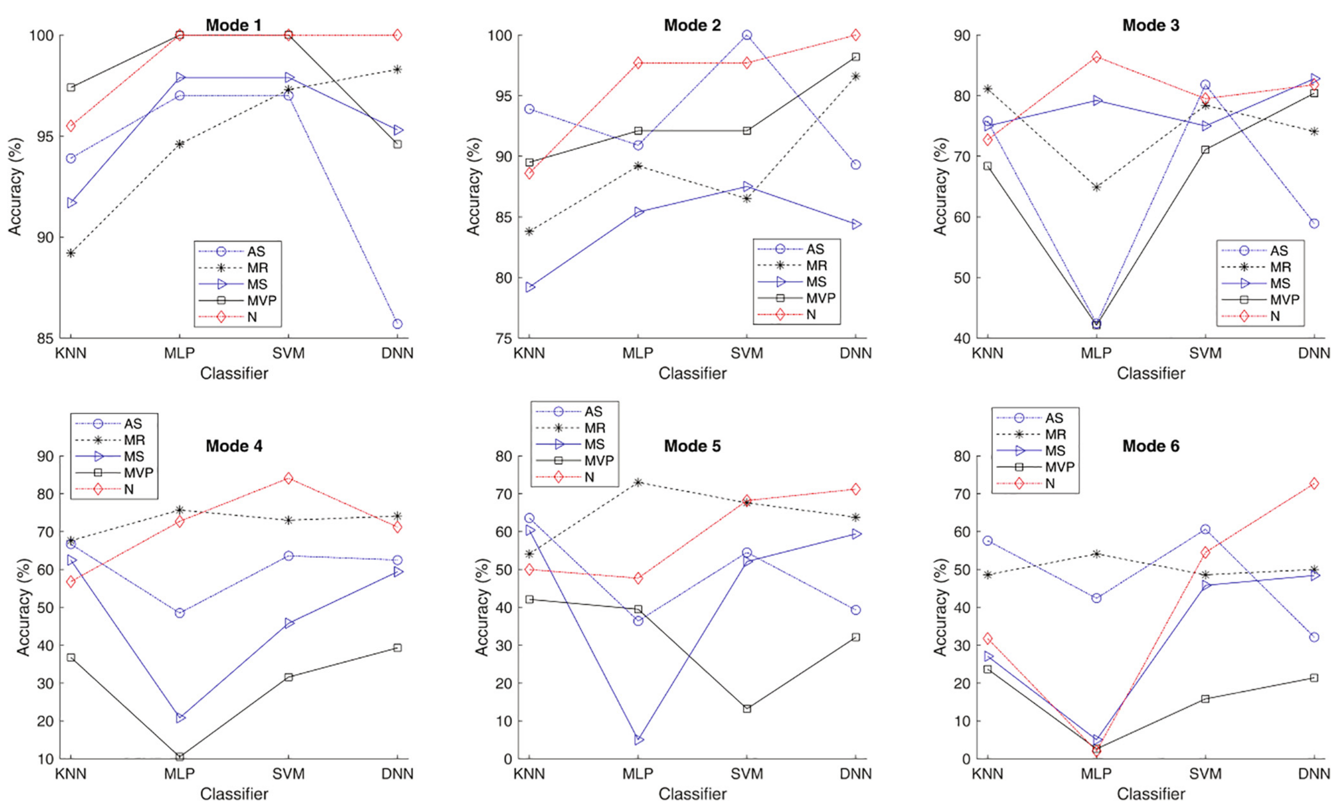

Fig. 9. Accuracy results of healthy (N) and diseased (AS, MR, MS and MVP) PCG signals by models for each mode.

classes. It is observed that DNN and SVM models reach higher values among the accuracy values of other models.

Table 2 summarizes the performances of the four models for all modes and the combined version of these modes. While MFCC features are used directly without feature selection for the DNN model, feature selection is performed with genetic algorithm in KNN, MLP and SVM models. For the dataset of size 1000*13 for each mode and 1000*78 for combined all modes, the genetic algorithm is applied, optimized feature subset is obtained, and results are compared. In all versions, the DNN model has provided the highest precision (Pr), recall (Rc), F1-score (F1) and accuracy (Acc) results compared to other models with five-fold cross validation.

As seen in Table 2, 89.1% accuracy is achieved when the DNN model is used with MFCC features, and 97.1% and 93.5% accuracy when used with MFCC features in mode 1 and mode 2, respectively. This result shows that the frequency components of healthy and pathological PCG signals are transferred to different modes with the decomposition process. The highest accuracy value of 98.9% is reached when the 78 (6x13) MFCC features obtained by combining all modes are used in the training and testing of the DNN model. The DNN model is followed by the SVM model with 96.2% accuracy. Also, it is clearly seen that quite low values are reached for all models in mode 3 and higher. Finally, the overall performance of all models degrades in mode 3 and higher. Experimental results show that the features obtained from the first two modes play a decisive role in the classification of healthy and diseased PCG sound signals.

The 10 $\times$ 5-fold CV DAP are applied to the EMD-based MFCC features obtained from PCG signals to compare the performance of KNN, SVM, MLP and DNN classifier models. The DAP results of all models for one healthy and four diseased PCG classes are given in Table 3. The performance measure in DAP internal validation is MCC with 95% studentized bootstrap confidence intervals (min CI, max CI). The total number of features is 78 (6 modes $\times$ 13 MFCCs) and the models are calculated for p = {25%, 50%, 75% and 100%} of the total number of features for each task. The DAP framework results given in Table 3 show that the DNN classifier, with an MCC value of 0.981 (min CI = 0.977 and max CI = 0.985), is more successful in categorizing PCG signals than other classifiers.

4. Discussion

The use of machine learning-based systems in disease detection of PCG sound signals has been the subject of research for a long time. Table 4 summarizes the studies in the literature on this subject. The binary classification of PCG signals as healthy and diseased is widely discussed. However, the development of new signal processing methods and technology enables the detection of the disease type. There are a limited number of multi-classification studies in the literature for the detection of disease types.

The objective of this study is to classify various kinds of cardiovascular abnormalities from PCG signals using HHT based analysis and classification approach. The MFCC features are extracted from the HHT modes to capture the frequency changes in the PCG signals. It has been observed that the mode 1 and 2 MFCC features are discriminative and these features have provided higher performance for classification of PCG signals with DNN model. The proposed study is compared with previous studies for automatic classification of PCG signals and the results are given in Table 4.

In recent years, in studies on binary classification for PCG signals, (Potes et al., 2016; Li et al., 2020a; Chen et al., 2020) have used time–frequency, high-order statistics and wavelet features with CNN classifier, respectively. Results of the studies show that the higher accuracy value of 94% have been obtained by (Chen et al., 2020) using the same dataset. The method reported by Raza et al. (2019) used frames from PCG signals and LSTM-RNN classifier for the detection of normal, murmur and systole. They have obtained overall accuracy of 80.45% using LSTM-RNN model. Similarly, in (Aziz et al., 2020), authors have extracted MFCCs and LTPs features and used the SVM model for the detection of normal, ASD and VSD signals. The classification of HVDs based Table Abbrevia-

Table 2
Overall performances of the classification models for all modes with 5-fold cross validation.

Features	Machine Learning Techniques															
	KNN				MLP				SVM				DNN			
	Pr	Rc	F1	Acc	Pr	Rc	F1	Acc	Pr	Rc	F1	Acc	Pr	Rc	F1	Acc
MFCCs	84.3	84.2	84.2	84.3	81.6	80.8	80.8	80.9	86.2	86.1	86.1	86.1	89.1	89.0	89.0	89.1
Mode-1 MFCCs	91.3	90.9	90.8	90.9	90.7	90.6	90.4	90.5	95.2	94.8	94.9	94.9	97.2	96.9	97.0	97.1
Mode-2 MFCCs	87.2	86.5	86.6	86.5	77.2	77.0	76.5	76.7	90.9	90.5	90.5	90.5	93.3	93.5	93.4	93.5
Mode-3 MFCCs	71.6	70.9	70.8	71.1	60.4	60.3	58.9	60.0	76.8	76.6	76.4	76.5	84.3	84.2	84.2	84.3
Mode-4 MFCCs	57.8	57.7	57.2	57.8	42.1	39.2	33.1	39.6	57.7	57.4	56.6	57.2	67.9	67.7	67.8	67.8
Mode-5 MFCCs	44.0	43.9	43.6	44.0	23.0	29.1	21.1	29.6	46.9	47.3	44.5	47.2	58.9	58.7	58.8	58.9
Mode-6 MFCCs	37.3	37.4	36.8	37.4	13.2	25.3	16.2	25.5	39.4	39.9	39.2	39.9	51.4	51.3	51.3	51.4
All Mode MFCCs	88.8	88.4	88.3	88.4	89.4	88.9	88.6	88.8	96.4	96.1	96.1	96.2	**98.9**	**98.7**	**98.8**	**98.9**

Table 3
Performance of classification models using DAP framework.

Features	KNN			MLP			SVM			DNN		
p	MCC	min CI	max CI	MCC	min CI	max CI	MCC	min CI	max CI	MCC	min CI	max CI
20	0.749	0.738	0.757	0.785	0.779	0.789	0.806	0.801	0.812	0.862	0.857	0.866
39	0.795	0.788	0.804	0.811	0.803	0.820	0.882	0.877	0.886	0.904	0.896	0.913
59	0.841	0.837	0.845	0.857	0.852	0.861	0.948	0.945	0.952	0.959	0.956	0.962
78	0.873	0.870	0.877	0.891	0.887	0.894	0.960	0.955	0.966	0.981	0.977	0.985

Table 4
Comparison of proposed method and related works.

Authors	Dataset	Feature extraction methods	Classifier	Acc. (%)
Potes et al. (2016)	PhysioNet/CinC	time-freq. features	CNN	89
Yaseen and Kwon (2018)	Own dataset	MFCCs and DWT features	SVM	97.9
Li et al. (2020a)	PhysioNet/CinC	high-order statistics	1D-CNN	86.8
Raza et al. (2019)	Pascal	signal frames	LSTM-RNN	80.45
Ghosh et al. (2019)	Yaseen and Kwon (2018) dataset	magnitude and phase features	RF	95.13
Chen et al. (2020)	PhysioNet/CinC	wavelet features	CNN	94
Aziz et al. (2020)	Own dataset	MFCCs and 1D-LTPs	SVM	95.24
Kumar et al. (2020)	Yaseen and Kwon (2018) dataset	Chirplet transform	Composite	98.33
He et al. (2021)	PhysioNet/CinC	Signal segmented on U-net	CNN	96.40
Zeng et al. (2021b)	Yaseen and Kwon (2018) dataset	Tunable Q-factor wavelet transform	Deep Wavelet	98.48
This study	Yaseen and Kwon (2018) dataset	MFCCs based on EMD	SVM	96.2
This study	Yaseen and Kwon (2018) dataset	MFCCs based on EMD	DNN	98.9

tions on the time–frequency analysis using Chirplet transform of the PCG signal has been reported by (Kumar et al., 2020). They have obtained an overall accuracy of 98.33% using multiclass composite classifier for detection of AS, MR, MS and healthy PCG signals. Yaseen and Kwon (2018) reported that they extracted MFCC and DWT features from PCG signals and used SVM and DNN models for the classification of AS, MS, MR, MVP and normal PCG signals. The results of this study were reported with accuracy of 91.6% when using MFCC features with SVM, and accuracy of 92.1% and 97.9% when combining both MFCC and DWT features with DNN and SVM models, respectively. Similarly, in (Ghosh et al., 2019), authors have extracted magnitude and phase features from PCG signals for the detection of heart abnormalities. They have reported an accuracy of 95.13% using a RF classifier. However, our proposed method has yielded higher performance compared to the methods outlined in Table 4. In this study, in addition to the traditional cross validation method, the 10x5-fold cross validation DAP method was used, which provides a strong estimation and minimizes the overfitting effect. According to DAP framework results, DNN classifier is superior to other classification models.

The originality and motivation of our proposed method are as follows. MFCC features are obtained from the signals decomposed into modes by performing time–frequency analysis with Hilbert-Huang transform. In the classification of PCG signals, the extracted features for each mode are used with the KNN, MLP, SVM and DNN models and the effects of the modes on the classification are analyzed. The proposed classification approach is effective in that it decomposes signals into modes according to frequency components. The proposed approach is based on the Hilbert-Huang transform, which uses the signal itself as a basis function, and is superior to other approaches based on the Fourier and wavelet transform in detecting pathological signals. Our proposed approach can be implemented in real-time systems in disease detection from PCG signals which can be recorded using the digital stethoscope.

5. Conclusions

In this work, classification of healthy and diseased PCG signals based on time–frequency analysis using the HHT has been proposed. The MFCC features have been computed and evaluated using decomposed modes of PCG signal. In detection disease of PCG signals, MFCC features extracted from PCG modes have been used with various classifiers such as KNN, MLP, SVM and DNN. Genetic algorithm has been used in feature selection for KNN, MLP and SVM models. In the training and testing of the DNN model, all features have been used without feature selection. The proposed Hilbert-Huang transform based feature extraction approach has shown better performance (overall accuracy of 98.9%) for classification of PCG signals using the DNN classifier. In addition, according to the performed the DAP method, the DNN classifier has MCC of 0.981 (min CI = 0.977 and max CI = 0.985) value and is more successful than other models. In the future, noise robust ensemble EMD and complementary ensemble EMD decomposition methods can be developed to extract features from the PCG signal.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Abbreviations

PCG	Phonocardiogram
HHT	Hilbert-Huang transform
EMD	Empirical mode decomposition
IMFs	Intrinsic mode functions
MFCC	Mel-frequency cepstral coefficient
GA	Genetic algorithm
ML	Machine learning
KNN	K-nearest neighbor
MLP	Multilayer perceptron
SVM	Support vector machine
DNN	Deep neural network
ReLU	Rectified linear unit
ADAM	Adaptive moment estimation

References

Alkhodari, M., Fraiwan, L., 2021. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. Comput. Methods Programs Biomed. 200, 105940.

Ari, S., Hembram, K., Saha, G., 2010. Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier. Expert Syst. Appl. 37 (12), 8019–8026.

Arslan, Ö., Engin, E.Z., 2019. Speech enhancement using adaptive thresholding based on gamma distribution of Teager energy operated intrinsic mode functions. Turk. J. Elec. Eng. Comp. Sci. 27 (2), 1355–1370.

Aziz, S., Khan, M.U., Alhaisoni, M., Akram, T., 2020. Phonocardiogram signal processing for automatic diagnosis of congenital heart disorders through fusion of temporal and cepstral features. Sensors 20 (13), 3790.

Babatunde, O.H., Armstrong, L., Diepeveen, D., Oluleye, B., Leisa, A., Dean, D., 2014. A genetic algorithm-based feature selection. Int. J. Electron. Commun. Comput. Eng. 5, 899–905.

Baydoun, M., Safatly, L., Ghaziri, H., El Hajj, A., 2020. Analysis of heart sound anomalies using ensemble learning. Biomed. Signal Process. Control 62, 102019.

Bishop, C.M., 2006. Pattern recognition. Machine Learning, Springer 128 (9).

Bozkurt, B., Germanakis, I., Stylianou, Y., 2018. A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection. Comput. Biol. Med. 100, 132–143.
Brereton, R.G., Lloyd, G.R., 2010. Support vector machines for classification and regression. Analyst 135 (2), 230–267.
Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., Xie, Z., 2018. Deep learning and its applications in biomedicine. Genomics Proteomics Bioinform. 16 (1), 17–32.
Chen, Y., Wei, S., Zhang, Y., 2020. Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network. Med. Biol. Eng. Compu. 58 (9), 2039–2047.
Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21 (1), 1–13.
Chicco, D., Warrens, M.J., Jurman, G., 2021. The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. IEEE Access 9, 78368–78381.
Chierici, M., Bussola, N., Marcolini, A., Francescatto, M., Zandonà, A., Trastulla, L., Agostinelli, C., Jurman, G., Furlanello, C., 2020. Integrative Network Fusion: a multi-omics approach in molecular profiling. Front. Oncol. 10, 1065.
Deng, M., Meng, T., Cao, J., Wang, S., Zhang, J., Fan, H., 2020. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Netw. 130, 22–32.
Deperlioglu, O., Kose, U., Gupta, D., Khanna, A., Sangaiah, A.K., 2020. Diagnosis of heart diseases by a secure internet of health things system based on autoencoder deep neural network. Comput. Commun. 162, 31–50.
Fioravanti, D., Giarratano, Y., Maggio, V., Agostinelli, C., Chierici, M., Jurman, G., Furlanello, C., 2018. Phylogenetic convolutional neural networks in metagenomics. BMC Bioinf. 19 (2), 1–13.
Francisco, S., 1959. A general theory of the causes of murmurs in the cardiovascular system. Am. J. Med. 27 (3), 360–374.
Fröhlich, H., Chapelle, O. and Schölkopf, B., 2003. Feature selection for support vector machines by means of genetic algorithms, in: Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence, IEEE, 2003. pp. 142-148.
Ghosh, S.K., Tripathy, R.K., Pachori, R.B., 2019. Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features. IEEE Sens. Lett. 3, 1–4.
Gonz, J., Phoo, C.P., Wiens, J., 2016. Heart sound classification based on temporal alignment techniques. In: IEEE 2016 Computing in Cardiology Conference; Vancouver, British Columbia, Canada. pp. 589–592.
He, Y., Li, W., Zhang, W., Zhang, S., Pi, X., Liu, H., 2021. Research on segmentation and classification of heart sound signals based on deep learning. Appl. Sci. 11 (2), 651.
Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intel. Systems Their Appl. 13 (4), 18–28.
Kramer, O., 2013. K-nearest neighbors, . pp.13–23.
Khaldi, K., 2012. Processing and analysis of sounds signals by Huang transform (Empirical Mode Decomposition: EMD). PhD, Université de Bretagne Occidentale, Bretagne, France.
Kim, M., 2013. Feature Extraction using MFCC. Signal Image Process. 4 (4), 101–108.
Kobat, M.A., Dogan, S., 2021. Novel three kernelled binary pattern feature extractor based automated PCG sound classification method. Appl. Acoust. 179, 108040.
Kumar, A.K., Saha, G., 2018. Improved computerized cardiac auscultation by discarding artifact contaminated PCG signal sub-sequence. Biomed. Signal Process. Control 41, 48–62.
Kumar, S., Ponnalagu, R.N., Tripathy, R.K., Acharya, U.R., 2020. Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals. Comput. Biol. Med. 118, 103632.
Laaksonen, J., Oja, E., 1996. Classification with learning k-nearest neighbors. in: Proceedings of International Conference on Neural Networks (ICNN'96), IEEE, 1996, pp.1480–1483.
Li, F., Tang, H., Shang, S., Mathiak, K., 2020a. Classification of heart sounds using convolutional neural network. Appl. Sci. 10 (11), 39–56.
Li, Z., Zhou, D., Wan, L., Li, J., Mou, W., 2020b. Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram. J. Electrocardiol. 58, 105–112.
Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Clifford, G.D., 2016. An open access database for the evaluation of heart sound algorithms. Physiol. Meas. 37 (12), 2181–2213.
Maggio, V., Chierici, M., Jurman, G., Furlanello, C., 2018. Distillation of the clinical algorithm improves prognosis by multi-task deep learning in high-risk neuroblastoma. PLoS ONE 13, (12) e0208924.
Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A., 2009. Support vectors machine-based identification of heart valve diseases using heart sounds. Comput. Methods Programs Biomed. 95 (1), 47–61.
Mavroforakis, M.E., Theodoridis, S., 2006. A geometric approach to support vector machine (SVM) classification. IEEE Trans. Neural Networks 17, 671–682.
Meintjes, A., Lowe, A., Legget, M., 2018. Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks. In: IEEE 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Honolulu. HI, USA, pp. 409–412.
Narendra, P.M., 1975. A branch and bound algorithm for computing k-nearest. IEEE Trans. Comput. 100 (7), 750–753.
Noman, F., Ting, C., Salleh, S., Ombao, H., 2019. Short-segment heart sound classification using an ensemble of deep convolutional neural networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing; Brighton, UK, pp. 1318–1322.
Potes, C., Parvaneh, S., Rahman, A., Conroy, B., 2016. Ensemble of feature-based and deep learning based classifiers for detection of abnormal heart sounds. In: IEEE 2016 Computing in Cardiology Conference; Vancouver, British Columbia, Canada, pp. 621–624.
Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G.S., On, B.W., 2019. Heartbeat sound signal classification using deep learning. Sensors. 19 (21), 1–15.
Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Commun. 54 (4), 543–565.
Saravanan, R., Sujatha, P., 2018. Algorithms : A Perspective of Supervised Learning Approaches in Data Classification, 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS). (2018) 945–949.
Shuvo, S.B., Ali, S.N., Swapnil, S.I., Al-Rakhami, M.S., Gumaei, A., 2021. CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. IEEE Access 9, 36955–36967.
Tan, F., Fu, X., Zhang, Y., Bourgeois, A.G., 2008. A genetic algorithm-based method for feature subset selection. Soft. Comput. 12 (2), 111–120.
Taud, H., Mas, J.F., 2018. Multilayer perceptron (MLP) neural networks, in: Geomatic Approaches for Modeling Land Change Scenarios. Springer, Cham, 2018, pp. 451–455.
Thiyagaraja, S.R., Dantu, R., Shrestha, P.L., Chitnis, A., Thompson, M.A., Anumandla, P.T., Dantu, S., 2018. A novel heart-mobile interface for detection and classification of heart sounds. Biomed. Signal Process. Control 45, 313–324.
Tschannen, M., Kramer, T., Marti, G., Heinzmann, M., Wiatowski, T., 2016. Heart sound classification using deep structured features. In: IEEE 2016 Computing in Cardiology Conference; Vancouver, British Columbia, Canada, pp. 565–568.
Yaseen, G.Y.S., Kwon, S., 2018. Classification of heart sound signal using multiple features. Appl. Sci. 8 (12), 23–44.
Zeng, W., Yuan, J., Yuan, C., Wang, Q., Liu, F., Wang, Y., 2021a. A new approach for the detection of abnormal heart sound signals using TQWT, VMD and neural networks. Artif. Intell. Rev. 54 (3), 1613–1647.
Zeng, W., Lin, Z., Yuan, C., Wang, Q., Liu, F., Wang, Y., 2021b. Detection of heart valve disorders from PCG signals using TQWT, FA-MVEMD, Shannon energy envelope and deterministic learning. Artif. Intell. Rev., 1–38
Zheng, F., Zhang, G., Song, Z., 2001. Comparison of different implementations of MFCC. J. Comput. Sci. Tech-Ch. 16 (6), 582–589.
Zhong, M., Yang, Y., Zhou, Y., Octavian Postolache, M., Chandrasekar, M., Venkat Babu, G., Elamaran, V., 2020. Advanced digital signal processing techniques on the classification of the heart sound signals. J. Med. Imaging Health Infor. 10 (9), 2010–2015.