Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2018) 26: 1555 – 1571

doi:10.3906/elk-1706-247

Bagged tree classification of arrhythmia using wavelets for denoising, compression, and feature extraction

Özgür TOMAK[1,*], **Temel KAYIKÇIOĞLU**[2]
[1]Department of Electric Power & Energy, Vocational School of Technical Sciences, Giresun University, Giresun, Turkey
[2]Department of Electrical & Electronics Engineering, Faculty of Engineering, Karadeniz Technical University, Trabzon, Turkey

Received: 21.06.2017 • **Accepted/Published Online:** 03.04.2018 • **Final Version:** 30.05.2018

Abstract: Arrhythmia, also known as dysrhythmia, is a condition involving an irregular heartbeat. A problem in the heart may cause problems in other organs, and as time passes, this will lead to more severe problems. Arrhythmia must be detected at an early stage to prevent such a problem occurring in the heart. Detection of arrhythmia from an electrocardiogram is an easy method that does not need much equipment and does not harm the patient. The purpose of this research is to find a faster and more accurate system to classify nine classes of arrhythmia. The St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database was used for training and testing. Data were compressed and preprocessed (denoising, trend elimination, baseline correction, and normalization) before being sent to the system for feature calculation. The wavelet coefficients that displayed the most significant effect on classification were chosen and used as features. Standard deviation and variance were also added to the feature set. Later, principal component analysis (PCA) was used to reduce the number of features further. After deciding the features, the performance of the basic classification methods and spiking neural network was checked to determine whether there was a better classifier to be used for our research. Tenfold cross-validation was applied to the training dataset. Bagged trees were found to produce better results. The classifiers' performance was tested by sensitivity, specificity, and accuracy.

Key words: Electrocardiogram, arrhythmia, wavelet, principal component analysis, bagged tree classification

1. Introduction

The increasing number of deaths due to arrhythmia has required automatic detection of abnormalities for identifying arrhythmias. Arrhythmias usually go unnoticed until a physician suspects the condition. Such an automatic detection system would help physicians to identify arrhythmias and reduce the pressure they experience when attending to such a severe ailment.

The detection of arrhythmia is a pattern recognition problem, and data obtained from electrocardiograms (ECGs) are of vital importance for detecting arrhythmia. By using electrodes that are placed on the patient, it is possible to measure and record the electrical activity of the heart. This record is called an ECG. The exact placement of 10 electrodes on the patient (an electrode for each leg and arm and six placed across the chest) is necessary for 12-lead ECG recording.

*Correspondence: ozgur.tomak@giresun.edu.tr

This research aimed to develop a fast and trustworthy system for the automated diagnosis of arrhythmia. To accurately extract features from ECG signals, we used the capability of the wavelet transform (WT) to process signals at different resolutions or scales.

The significance of this research is exhibited in several factors. The first is the usage of wavelet-compressed data from the St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database. The second is the usage of two small windows when calculating wavelet coefficients. The third is the significantly reduced number of features. This reduction was achieved by determining coefficients that have the most significant effect on classification and the usage of principal component analysis (PCA), and the fourth is the usage of bagged trees (ensemble classifier) for classification to compare the performance of the 12-lead ECG recordings. The usage of bagged trees was chosen after a comparison of the performance of the basic classification methods and spiking neural network. The data were randomized, and half of the data were used for training. The other half of the data were used as test data for obtaining classification sensitivity, specificity, and accuracy values.

2. Related works

Dokur et al. [1] investigated wavelet analysis and Fourier analysis as feature extraction methods. Ten different ECG waveforms were classified using a neural network trained by genetic algorithms. The proposed method provided performance with 99.4% accuracy using WT and 92.2% accuracy using discrete Fourier transform.

Prasad and Sahambi [2] analyzed ECG arrhythmias using wavelets. All coefficients obtained from discrete wavelet transform (DWT) were used as features. They preferred artificial neural networks (ANNs) for the classification of 12 different arrhythmias. An accuracy rate of 96.77% was obtained. This method was found to be robust against noise.

Shyu et al. [3] developed a method for classifying ventricular premature contraction. QRS detection and feature extraction were done using WT. Less complexity and less computation were the primary advantages of this technique. The accuracy of ventricular premature contraction classification using a fuzzy neural network was 99.79%.

Zhao and Zhang [4] used WT to extract coefficients. Simultaneously, they applied autoregressive modeling so that temporal structures of ECG waveforms were obtained. Then a support vector machine (SVM) with Gaussian kernel was used to classify six heart rhythm types. Overall accuracy of 99.68% was achieved.

Jiang et al. [5] proposed a feature extraction method from the combination of WT and independent component analysis (ICA). The SVM classifier was used to recognize fourteen different heartbeat arrhythmias, and they obtained overall accuracy of 98.65%.

Yu and Chen [6] proposed a system to classify six ECG beat types. This system used DWT to decompose the ECG signals in different subbands. They used a probabilistic neural network for the classification. An accuracy rate of 99.65% was obtained with 11 features. This feature set was significantly smaller in quantity than the feature sets of other methods.

İşler and Kuntalp [7] used a genetic algorithm to select the best features from heart rate variability measures and wavelet entropy measures. The MIT-BIH database was preferred. The K-nearest neighbor classifier was used to classify two groups (congestive heart failure and healthy). The best accuracy was 96.39%.

An extremely accurate ECG beat detection system was proposed by Zellmer et al. [8]. Three separate feature vectors for each beat were formed. Continuous wavelet transformation (CWT) in addition to time domain morphology information was necessary to create these feature vectors. Three different SVM classifiers were trained by these feature vectors for six types of beats. They obtained an accuracy rate of 99.72% using multiclassifier-based classification by voting between the three independent classifiers.

Emanet [9] classified the five types of beats using DWT. The random forest algorithm was used for classification. Twenty-three records from the MIT-BIH arrhythmia database were randomly chosen. Moreover, five heartbeat types were classified with a 99.8% success rate.

Llamedo and Martínez [10] used features from the RR signals and different scales of the WT. The best model was trained in a partition of the MIT-BIH arrhythmia database. The classification was done by a linear classifier, and global accuracy of 93% was obtained.

Rai et al. [11] conducted a study on ECG arrhythmia classification using a multilayer perceptron and radial basis function neural network with Daubechies wavelets. ECG beats were characterized by 21 points for five types of beats. The simulation resulted in an average accuracy rate of 99.84% with an error rate of 0.16%.

Rai and Trivedi [12] classified ECG signals into two classes using DWT-based features and morphological features. They used a backpropagation neural network for classification and obtained an accuracy rate of 97.8%.

Li et al. [13] utilized ECG signals from the Research Resource for Complex Physiologic Signals for detecting ventricular fibrillation. CWT was used for feature extraction. This method could detect ventricular fibrillation without interrupting the ongoing chest compression. Morphology consistency evaluation obtained results with an accuracy of 93%, sensitivity of 92%, and specificity of 93%.

Ye et al. [14] proposed an approach for classification based on a combination of dynamic and morphological features. The MIT-BIH arrhythmia database was used. WT and also ICA were applied to signals to obtain morphological features. Exact locations of RR intervals were computed to provide dynamic features. These features were concatenated and classified using SVM. The overall accuracy was 99.3%.

Llamedo and Martínez [15] developed a patient-adaptable algorithm. Morphology descriptors computed from the WT and RR interval features were used. The American Heart Association database and databases available from PhysioNet were used to classify normal, supraventricular, and ventricular heartbeats. An automatic classifier performed cluster and centroid identification. MITBIH-SUP and the subset of MITBIH-AR were used for the final performance evaluation, and 96% accuracy was achieved.

Faziludeen and Sabiq [16] implemented an automatic classification method for three beat types. Twenty-five features were extracted from wavelet analysis for each beat. Three SVMs were designed for classification. Final grouping was done by maximum voting. Accuracy of 99.92%, 98.47%, and 98.46% was obtained for premature ventricular contraction, normal, and left bundle branch block beats, respectively.

Chen et al. [17] developed a system for classification of ECG signals based on linear discriminant analysis. Linear discriminant analysis was preferred for its simplicity. Spectral energy of ECG signals was used as features by obtaining the DWT of the PQRST complex. The Haar wavelet was used as the basis function. The PTB database and samples from the Southampton General Hospital Cardiology Department were used for classification, and 85.57% accuracy was achieved.

Dima et al. [18] developed a system for detection of the presence of a myocardial scar from ECGs and vector cardiograms. Four distinct methods were implemented. One of them was the usage of a template ECG heartbeat, coupled with wavelet coherence analysis. They used the SVM for classification. They also employed feature selection to remove redundant features. A total of 260 records from 3 different databases were used from the Cardiology Department of the University Hospital Southampton, and 89.22% accuracy was achieved.

Saminu et al. [19] calculated the DWT of the R-T interval and the statistical parameters were used as time-frequency domain features. Three types of ECG beats were classified using a neural network backpropagation algorithm. They obtained an accuracy rate of 98.22%. This technique is a robust ECG feature extraction method that is suitable for mobile devices.

Banerjee and Mitra [20] used cross-wavelet transform (XWT) for classification of ECG signals. They analyzed ECG signals utilizing XWT and explored the resulting spectral differences. They also obtained the wavelet coherence (WCOH) and wavelet cross-spectrum (WCS) by applying XWT to signals. The WCOH and WCS of ECG signals showed different characteristics over two specific parts (QRS complex and T wave part). The classifier was preferred as being threshold-based. The overall accuracy, specificity, and sensitivity were obtained as 97.6%, 98.8%, and 97.3%, respectively, for classification of abnormal and normal cardiac patterns.

Sahoo et al. [21] planned to detect cardiac arrhythmia from ECG signals by using WT. The first 23 files from the MIT-BIH arrhythmia database were used. In this work, five abnormal and five normal ECG signals were chosen for analysis. A least squares SVM was used for classification. They obtained 98.11% accuracy.

Jacob and Joseph [22] performed classification using XWT and SVMs for two classes. The simulation resulted in an accuracy rate of 94.8% for the SVM and 96.2% for the two-dimensional SVM.

Saini et al. [23] classified ten heart diseases using the K-nearest neighbors (KNN) classifier. The classification was done using wavelet-transformed ECG signals and the original ECG signals. An accuracy rate of 87.5% was obtained.

Sharma et al. [24] used eigenvalues of multiscale covariance matrices and multiscale wavelet energies as features. They used SVMs with both linear and radial basis function kernels and KNN as classifiers. The PTB diagnostic ECG database was used for classification. Datasets included various types of myocardial infarctions. The accuracy, the specificity, and the sensitivity values were 96%, 99%, and 93%, respectively.

Dewangan and Shukla [25] analyzed ECG signals using DWT for feature extraction and preprocessing. An ANN was used for the classification of five types of arrhythmias. They determined that the utilization of both wavelet coefficients and morphological features together for classification could increase the accuracy rate. The proposed method provided an enhanced performance with 87% accuracy.

Sayilgan et al. [26] used clustering algorithms to determine seven types of arrhythmia. They examined data from the MIT-BIH arrhythmia database. K-means, fuzzy C-means, an extreme learning machine, and naive Bayes were used as classifiers. Naive Bayes was detected as the most successful classifier with 92% accuracy.

Rad et al. [27] developed algorithms for automatic classification of ventricular tachycardia, ventricular fibrillation, pulseless electrical activity, asystole, and pulse generating rhythm. They used a database with 1631 three-second ECG signals. These signals were collected from 298 cardiac arrest patients (OHCA database). They computed 47 wavelet and time domain-based features. A wrapper-based feature selection architecture was used for selecting 14 features. An ANN classifier with Bayesian regularization was used. The overall accuracy was 78.5%.

3. Wavelet theory

Wavelet theory is used as a powerful mathematical tool. Signals are decomposed over translated and dilated wavelets by WT.

3.1. Wavelet families

The mother wavelet (analyzing wavelet) is the prototype function adapted to wavelet analysis [28]. The known wavelet families are as follows: Haar, Daubechies, Dymlets, Coiflets, biorthogonal, Fejer–Korovkin filters, reverse biorthogonal, Meyer, discrete approximation of Meyer, Gaussian, Mexican hat, Morlet, complex Gaussian, Shannon, frequency B-spline, and complex Morlet. These are the wavelets available within the wavelet toolbox.

Daubechies wavelets are found to be the most useful for analyzing ECG signals after checking these wavelets. The Daubechies wavelet transforms produce differences and averages using a few more values from the ECG signal. Moreover, this change provides vast improvement in the capabilities of these transforms. Powerful tools for signal recognition are provided [29]. Its shape is similar to the ECG signal.

3.2. Wavelet transform

Wavelet function $\Psi\gamma L^2(\mathrm{R})$ is zero-averaged, normalized, and centered on $t = 0$ where f is the signal that is transformed [30].

We can write WT of $f\gamma L^2(\mathrm{R})$ at time u and scale s as a convolution product:

$$W\,f\;(u\,s) \;=\; \int_{-\infty}^{+\infty} f\,(t)\,\frac{1}{\sqrt{s}}\,\Psi^*\left(\frac{t-u}{s}\right)\,dt \;=\; f\,\star\,\bar{\Psi}_s\,(u) \tag{1}$$

with

$$\bar{\Psi}_s\;(t) \;=\; \frac{1}{\sqrt{s}}\,\Psi^*\left(\frac{-t}{s}\right). \tag{2}$$

3.3. Discrete wavelet transform

Our orthogonal wavelet basis function is as follows:

$$\Phi_{(s,l)}\;(x) \;=\; 2^{\frac{-s}{2}}\,\Phi\,(2^{-s}\,x-l), \tag{3}$$

where l and s are variables that dilate and scale the mother function Φ to generate a family of discrete wavelets. Location index l gives position, and scale index s gives the wavelet's width [28].

The analyzing wavelet Φ is used in a scaling equation to span the data domain at different resolutions:

$$W\,(x) \;=\; \sum_{k\,=\,-1}^{N-2} (-1)^k\,c_{k+1}\,\Phi\,(2x+k), \tag{4}$$

where $W(x)$ is the scaling function and c_k are the wavelet coefficients. The wavelet coefficients must satisfy

$$\sum_{k=0}^{N-1} c_k = 2 \quad and \quad \sum_{k=0}^{N-1} c_k c_l \;=\; 2\delta_{l,0}, \tag{5}$$

where l is the location index and δ is the delta function [28].

The signal is split into high and low frequency components in the first level. This first low frequency subband that contains most of the energy is subsampled and decomposed again into high and low frequency subbands. This can be continued into K levels [31]:

$$y_{high}\,[k] \;=\; \sum_n x\,[n]\;g\,[2k-n], \tag{6}$$

$$y_{low}\,[k] \;=\; \sum_n x\,[n]\;h\,[2k-n] \tag{7}$$

where $h[N-1-n] = (-1)^n\,g(n)$ and N is the number of current samples of $x[n]$.

3.4. Wavelet-based compression

ECG signal compression techniques are becoming increasingly important because they make it easier to work with long duration signals in large databases. Furthermore, they simplify wireless transmission of ECG signals. In the case of heart diseases, ECG data can be transmitted from rural areas to hospitals to be inspected by expert cardiologists using telemedicine. After analyzing the data, medical suggestions can be sent to semiskilled persons in the countryside [32].

We decided to use compressed signals for faster processing. We compressed each signal and saved them as .mat files for feature extraction. After the coefficients were calculated, a specific threshold value was determined. Coefficients smaller than this threshold were ignored. Compressed signals consume fewer computer resources than the original signals when recorded or transmitted over the network. If the inverse transform is applied, an approximate signal similar to the original one is attained. Obtaining a similar signal means that most of the features of the original signals can be stored in the compressed signals. Signals were decomposed at level 5, and a Daubechies 6 wavelet was used as the mother wavelet. For threshold criteria, global thresholding, which deploys hard thresholding, was used as the thresholding method. Balance sparsity-norm was selected for the global threshold. We can conclude by looking at Table 1, where we see that the classification performance does not change much if compressed data are used. PCA was used to reduce the selected features by deriving the variables that conserved most of the information. We achieved this by transforming a new set of variables, which were uncorrelated, to the principal components. Most of the variation was retained in the first few variables [33].

4. Materials and methods

Signals can be represented by using coefficients in a linear combination of wavelet functions. It is possible to use only these coefficients to perform operations. Furthermore, to compress data, one can truncate coefficients below a threshold. The signal was reduced in size via wavelet-based compression, making the signal easier to process. Each signal was compressed and saved as a .mat file using the MATLAB wavelet toolbox. These files were used for feature extraction. However, before feature extraction, preprocessing steps (denoising, trend elimination, baseline correction, and normalization) were executed for better results. Preprocessing steps are given in Figure 1. For denoising, a Daubechies wavelet (db6) was selected and the signals were decomposed at level 5. A soft fixed-form threshold was selected as the thresholding method, and the noise structure was selected as unscaled white noise. After denoising, the synthesized signal was updated. This function is available in the MATLAB wavelet toolbox. When we load the ECG signal, patterns can be seen that are not inherent to the ECG signal due to disturbances. Data analysis can be hindered by these trends. The linear trend was eliminated by using the function detrend, which is available in MATLAB. The nonlinear trend was eliminated by fitting a low-order polynomial to the ECG signal and subtracting it. Eq. (8) is for baseline correction, where the mean of the signal is subtracted from each value of the raw signal. Eq. (9) is for normalization, where each value signal is divided by the absolute value of its maximum value.

$$ECG\,Signal_{base\,line\,correction} = ECG\,Signal - mean(ECG\,Signal) \tag{8}$$

$$ECG\,Signal_{normalized} = \frac{ECG\,Signal}{|ECG\,Signal_{\max}|} \tag{9}$$

For each record, 60 samples were taken from either side of the R peak (maximum value). We used two windows, resulting in a total of 120 samples. This usage is different from most research, which tends to use only one window for feature extraction.

Table 1. Results obtained using the coefficients obtained from the 5th level decomposition of channel 1 with db1 / results obtained using compressed data by using the coefficients obtained from the 5th level decomposition of channel 1 with db1.

Classification methods	Training	Test		
	Accuracy	Accuracy	Sensitivity	Specificity
Ensemble classifiers				
Bagged trees	95.1 / 94.8	95.1 / 91.7	66 / 48	99 / 99
Boosted trees	93.3 / 91.5	93.4 / 89.6	57 / 20	98/ 99
Subspace discriminant	87.4 / 89.2	87.4 / 88.2	3 / 0	99 / 100
Subspace KNN	94.3 / 93.1	94.3 / 90.5	56 / 36	99 / 98
RUSBoosted trees	84.9 / 83.8	85.1 / 80.2	83 / 56	85 / 90
Nearest neighbor				
Fine KNN	93.7 / 91.1	93.93 / 88.1	74 / 68	97 / 96
Medium KNN	93.9 / 92.4	94.1 / 90.7	62 / 48	99 / 98
Coarse KNN	92.1 / 90.9	92.4 / 88.1	51 / 41	99 / 97
Cosine KNN	94.2 / 92.2	94.5 / 87.5	66 / 31	99 / 96
Cubic KNN	93.8 / 92.3	93.9 / 90.1	61 / 51	99 / 99
Weighted KNN	94.4 / 93.2	94.6 / 90.5	67 / 34	99 / 96
Support vector machines				
Linear SVM	87.2 / 89.2	87.3 / 88.3	0 / 0	100 / 100
Quadratic SVM	89.2 / 90.6	89.3 / 89.6	22 / 23	99 / 99
Cubic SVM	88.1 / 90.2	87.3 / 88.2	0 / 0	99 / 100
Fine Gaussian	93.6 / 92.2	93.9 / 91.1	54 / 43	99 /99
Medium Gaussian	92.7 / 91.6	92.9 / 90.8	52 / 44	99 / 99
Coarse Gaussian	88.4 / 89.6	88.6 / 88.2	13 / 5	99 / 99
Decision trees				
Simple tree	88.9 / 89.6	89.1 / 84.1	26 / 2	95 / 98
Medium tree	90.7 / 90.5	90.9 / 88.2	34 / 25	99 / 94
Complex tree	92.4 / 91.5	92.5 / 89.6	57 / 24	98 / 98
Discriminant analysis				
Linear discriminant	88.4 / 89.2	88.6 / 87.4	16 / 1	99 / 99
Quadratic discriminant	81.9 / 81.8	81.8 / 80.1	29 / 23	90 / 90

We decided not to use all the wavelet coefficients. The coefficients that had the most significant effect on classification were chosen and used as features. It is possible to classify much faster with fewer coefficients. For this to be achieved, first we checked classification performance using the first three coefficients. Then a fourth was added to observe its effect on classification. After this, we removed the first coefficient to observe its effect. Then a fifth was added; later, we removed the second. This technique was carried out repeatedly. We also added standard deviation and variance to the feature set. After forming feature vectors, the number of features was further reduced with PCA.

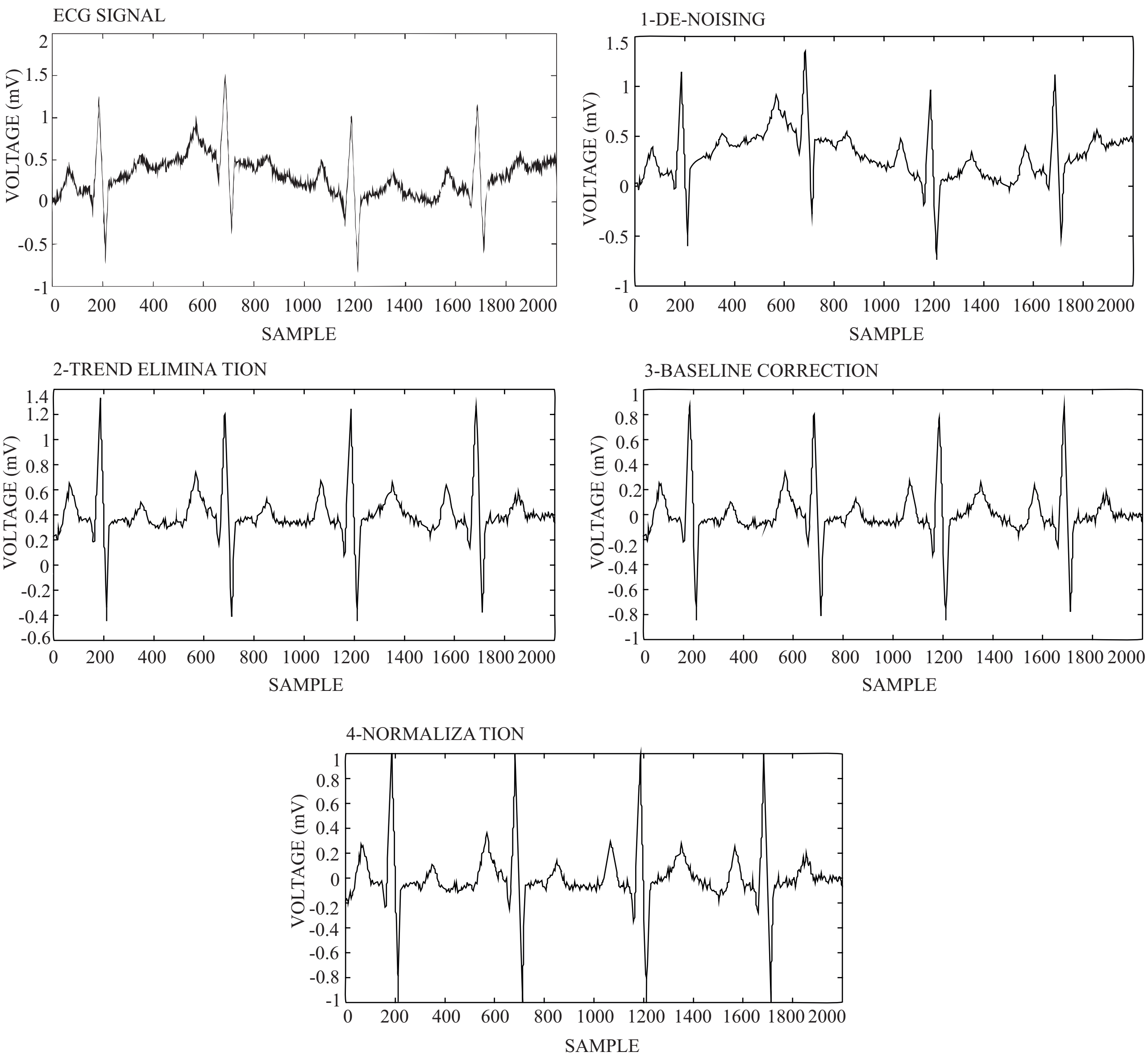

Figure 1. Preprocessing steps (denoising, trend elimination, baseline correction, and normalization).

5. Database

Most research uses the MIT-BIH arrhythmia database, which contains 48 half-hour, two-channel ECG recordings. However, we used the ECG signals that were collected in the St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database, which contains 75 half-hour annotated recordings. Each record is sampled at 257 Hz. These files belong to patients (15 women and 17 men, aged 18–80 years) with ECGs consistent with ischemia, conduction abnormalities, coronary artery disease, and arrhythmias [34].

Our first reason for choosing this database is that it contains 12-channel recordings. The MIT-BIH arrhythmia database contains only two-channel recordings. The 12-channel ECG records allow us to see the heart's electrical activity from 12 different perspectives. The second reason for choosing this database is that it is a big database. This database is more extensive than the MIT-BIH arrhythmia database. Testing our method on such a big database makes our results more reliable. Another reason for choosing this database is that subjects with ECGs consistent with ischemia, conduction abnormalities, coronary artery disease, and arrhythmias were included. Also, it should be noted that this database is a new one compared to the MIT-BIH arrhythmia database. That means that patients' histories are up to date.

6. Basic classification methods

Signals were classified by using data from the St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database. Half of the data were used for training. The remaining data were used for tests. In our research, there were nine cases for classification. Classes that were classified were as follows: normal (healthy), premature ventricular contraction, atrial premature beat, fusion of ventricular and normal beat, right bundle branch block beat, bundle branch block beat (unspecified), supraventricular premature or ectopic beat (atrial or nodal), nodal (junctional) escape beat, and unclassifiable beat. Classification Learner was often used, but sometimes MATLAB functions were preferred.

We decided on the classification method using Table 2. The accuracy, sensitivity, and specificity results obtained by using the reduced coefficients obtained from the 5th level decomposition of channel 1 with db6 are given in Table 2. Bagged trees were found to produce better results.

Bagging (bootstrap aggregation) was the ensemble learning technique that we chose for classification. The block diagram of the bagged decision tree is given in Figure 2 [35]. For bagged trees, the ensemble method was chosen as the bag. The learner type was a decision tree, which is easy to interpret. Moreover, the memory usage was low, as well. The maximum number of splits was set to 20, and the number of learners was set to 30. Tenfold cross-validation was applied to the training dataset. That means data were broken into ten sets. Nine datasets were trained and the remaining one dataset was used for the test. The process was repeated ten times and mean accuracy was taken. This cross-validation method protects data against overfitting.

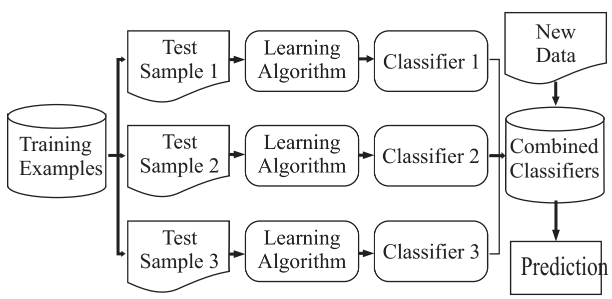

Figure 2. The block diagram of bagged decision trees [35].

Weak prediction models are ensembled to produce a better prediction model for boosted trees. The ensemble method chosen was AdaBoost, the decision tree was selected as learner type, the maximum number of splits was set to 20, the number of learners was set to 30, and the learning rate was set to 0.1. For the subspace discriminant, the ensemble method chosen was the subspace, and the discriminant was selected as the learner type. The number of learners was set to 30 and the subspace dimension was set to 2. For subspace KNN, the ensemble method chosen was the subspace and nearest neighbors were selected as the learner type. The number of learners was set to 30 and the subspace dimension was set to 2. For RUSBoosted trees, the ensemble method chosen was RUSBoost, and the decision tree was selected as the learner type. The maximum number of splits was set to 20, the number of learners was set to 30, and the learning rate was set to 0.1.

Table 2. The results obtained by using the reduced coefficients obtained from the 5th level decomposition of channel 1 with db6.

Classification methods	Training	Test		
	Accuracy	Accuracy	Sensitivity	Specificity
Ensemble classifiers				
Bagged trees	98.9	97.8	87	99
Boosted trees	93.2	93.0	51	99
Subspace discriminant	88.9	88.1	9	99
Subspace KNN	95.6	94.9	67	99
RUSBoosted trees	86.6	84.7	75	86
Nearest neighbor				
Fine KNN	94.2	93.8	68	98
Medium KNN	94.0	93.7	55	99
Coarse KNN	92.5	91.8	39	99
Cosine KNN	93.8	93.1	50	99
Cubic KNN	93.9	93.5	52	99
Weighted KNN	94.8	94.4	60	99
Support vector machines				
Linear SVM	88.1	87.5	1	99
Quadratic SVM	90.9	90.1	27	99
Cubic SVM	91.5	90.9	40	98
Fine Gaussian SVM	94.5	93.9	54	99
Medium Gaussian SVM	92.5	91.9	39	99
Coarse Gaussian SVM	88.7	88.1	6	99
Decision trees				
Simple tree	90.0	89.3	38	0.97
Medium tree	92.0	91.5	46	0.98
Complex tree	93.6	92.8	52	0.99
Discriminant analysis				
Linear discriminant	81.1	80.9	57	85
Quadratic discriminant	78.9	77.7	60	81

KNN classification methods were also tested because distance-based classification is an effective and straightforward method. Euclidean distance was chosen as the distance metric. We set the number of neighbors to 1 for fine KNN; 10 for medium KNN, cosine KNN, cubic KNN, and weighted KNN; and 100 for coarse KNN. The equal weighting function (no weights) was chosen as the distance weighting function for simplicity. Scaling each coordinate distance was another attribute to decide. For our research, we chose to standardize data.

We also tested SVMs, which find the best hyperplane (with the largest margin between classes) to separate data. By specifying the box constraint level, we can keep values of Lagrange multipliers in a bounded region. The box constraint level chosen was 1 for SVMs. Increasing the level can increase training time, but can also reduce the number of support vectors. The kernel scale was chosen as automatic for linear, quadratic, and cubic SVM; therefore, to select the scale value, a heuristic procedure was used. The kernel function was chosen as Gaussian for medium Gaussian, fine Gaussian, and coarse Gaussian SVMs.

Decision trees were also tested. Decision trees are one of the best choices for fast classification and are easy to interpret. A decision tree is a flowchart structure. The outcome of the test is represented by each branch. The class label is represented by each leaf node. Gini's diversity index was chosen as the split criterion. Employing a maximum number of splits was our way to control the depth; 4 was chosen for the simple tree classification, 20 for medium tree classification, and 100 for complex tree classification. For the discriminant analysis case, the selected regularization was diagonal covariance, which chooses the maximal possible regularization.

To evaluate the system, we ran it ten times, each time randomly dividing the data into a training set and a testing set. Classification accuracy, sensitivity, and specificity were calculated using true positive (TP, correctly identified events), true negative (TN, correctly rejected events), false positive (FP, incorrectly detected events), and false negative (FN, incorrectly rejected events) [27]. They are defined as:

$$Sensitivity = \frac{TP}{TP+FN} \times 100, \tag{10}$$

$$Specificity = \frac{TN}{TN+FP} \times 100, \tag{11}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100. \tag{12}$$

7. Deep learning

The convolutional neural network is one of the most successful classification techniques. In a simple convolutional neural network, the signals pass through a series of layers. Data are given to the system through an input layer. A set of filters is used in the convolution layer, and each filter is convolved across the signal. Dot products between the inputs and the entries of the filter are computed to produce an activation map. The layer of rectified linear units (ReLUs) is an activation function, which is a ramp function. This layer changes negative input values to zero. Nonlinear properties of the network are increased in this layer, and the system learns faster with the use of this layer. A pooling layer is usually placed after the ReLU layer. The primary goal of the pooling layer is to reduce the input size for the next convolution layer. It prevents the system from memorizing and creates a smaller computational load for the subsequent network layers. At this layer, the signal is passed through a filter according to a specific stepping value and is processed by taking the maximum or the average of the values. The fully connected layer connects each neuron in one layer to another neuron in another layer. The underlying logic implemented is to remove some nodes of the network in a dropout layer. The main goal of the dropout layer is to prevent the system from memorizing. A Softmax classifier is usually preferred in deep learning networks. The convolutional neural networks represent the state of the art for many classification problems [36].

Due to computational cost, convolutional neural networks are not suited for real-time applications. We therefore used spiking neural networks for faster calculations. It is a similar network designed for real-time applications. The spiking deep neural network was built as a realistic neural network to emulate the brain. It was found to be more efficient regarding computation at the expense of performance loss [36].

In spiking neural networks, inputs are presented as streams of events. The evidence is integrated by neurons during the presentation and spikes are created to communicate information. In our case, we used the wavelet coefficients as the input.

The integrate-and-fire (IF) model was used as the spiking neuron. Negative values in spiking neurons were avoided by using ReLUs. The weights from the ReLUs were directly mapped to a network of IF units. The bias was fixed to zero throughout training with backpropagation. Weight normalization was used to obtain faster convergence and near-lossless accuracy. The network weights were normalized to ensure that activations were small enough to prevent the ReLUs from overestimating output activations. All the weights were rescaled by the maximum possible positive input. A five-layer fully connected neural network was trained. We used a learning rate of 1 and 50% dropout. The weights were randomly initialized between 0.1 and –0.1. We obtained a training accuracy rate of 96.73% for the networks. We obtained an accuracy rate of 95.17% for the normalized network and an accuracy rate 95.22% for the unnormalized network on the test set. The normalization improved latencies without loss in accuracy. The accuracy of the normalized and unnormalized spiking neural network with respect to time is given in Figure 3.

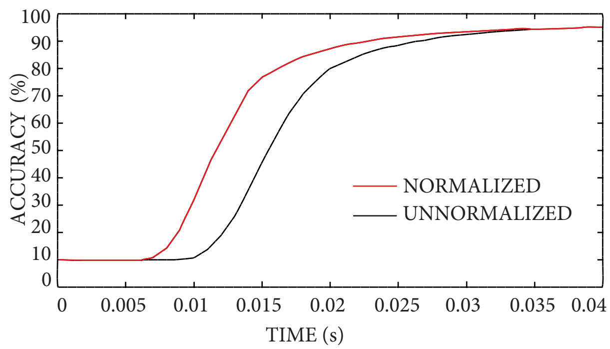

Figure 3. The accuracy of normalized and unnormalized spiking neural network.

8. Results

The average size of a 12-channel, 30-min signal was 11,102,420 bytes (10.6 MB). After compression, the average size of a signal for one channel was found to be 95.120 bytes (92.9 KB). It can be said that the file size drops to 1% of its original size. Around 99% of the energy was retained, which means almost all the energy was preserved (zeros 96.87%), and it is possible to reduce the amount of data considerably. Also, 2.23 s was needed to complete the compression of 30-min signals. After compressing the data, the time needed for calculations was reduced. Thus, the system is faster even though some time is needed for compression. The number of features was reduced to 6 by deciding on coefficients that had the most significant effect on classification and usage of PCA. Our system achieved 98.6% accuracy for training and 97.8% accuracy, 87.4% sensitivity, and 99.7% specificity for testing (average values). By looking at Table 3, we can say that a similar success rate was observed compared to the related works, but we tested our method on a bigger database. It took 777.626 s (average time) for the feature calculation of 75 half-hour ECG recordings with MATLAB on a computer with AMD A6-6310 processor and 4 GB RAM. After reduction of features, only 7.754 s was needed to complete the test. That means our system is faster and more stable.

9. Discussion

A quick and accurate diagnosis is possible using clinical decision support systems. These are biological signal processing applications created to help physicians. Also, using these applications, clinical management of heart diseases can be improved.

Table 3. Comparison of the related works with the proposed method.

Reference	Type	Channel	Database	Classifier	Accuracy
Dokur et al. [1]	10	1	MIT-BIH	Neural network	99.4
Prasad and Sahambi [2]	12	1	MIT-BIH	Neural network	96.77
Shyu et al. [3]	2	1	MIT-BIH	Neural network	99.79
Zhao and Zhang [4]	6	1	MIT-BIH	SVM	99.68
Jiang et al. [5]	14	1	MIT-BIH	SVM	98.65
Yu and Chen [6]	6	1	MIT-BIH	Neural network	99.65
İşler and Kuntalp [7]	2	1	MIT-BIH	KNN	96.39
Zellmer et al. [8]	6	1	MIT-BIH	SVM	99.72
Emanet [9]	5	1	MIT-BIH	Random Forest	99.8
Llamedo and Martínez [10]	2	1	MIT-BIH	Linear classifier	93
Rai et al. [11]	5	1	MIT-BIH	Neural network	99.84
Rai and Trivedi [12]	2	1	MIT-BIH	Neural network	97.8
Li et al. [13]	2	1	Complex Phys.	Morphology c.	93
Ye et al. [14]	16	1	MIT-BIH	SVM	99.3
Llamedo and Martínez [15]	3	1	PhysioNet	Cluster centroid	96
Faziludeen and Sabiq [16]	3	1	MIT-BIH	SVM	98.46
Chen et al. [17]	2	1	PTB database	Linear discrim.	85.57
Dima et al. [18]	2	1	Southampton	SVM	89.22
Saminu et al. [19]	3	1	MIT-BIH	Neural network	98.22
Banerjee and Mitra [20]	2	3	Physikalisch T.	Threshold-based	97.6
Sahoo et al. [21]	5	1	MIT-BIH	SVM	98.11
Jacob and Joseph [22]	2	1	MIT-BIH	SVM	94.8
Saini et al. [23]	10	1	MIT-BIH	KNN	87.5
Sharma et al. [24]	2	1	PTB Diag.	SVM	96
Dewangan and Shukla [25]	5	1	MIT-BIH	Neural network	87
Sayilgan et al. [26]	7	1	MIT-BIH	Naive Bayes	92
Rad et al. [27]	5	1	OHCA	Neural network	78.5
Proposed Method	9	12	St. Petersburg	Bagged tree	98.6

It is most probable that biomedical signal processing will play a key role in patient care at a distance [37]. When biomedical signals are transmitted via network systems, signal compression and fast, reliable analysis applications will be needed. When executing such work, efficiency and speed is a top priority.

The usage of wavelet-compressed data from the St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database was our first step. In general, the MIT-BIH arrhythmia database is used by most research that is performed on ECG signals [1–12,14,16,19,21–23,25,26], and classification is executed on a single channel [1–19,21–27]. Our research was performed on 12 channels. Furthermore, wavelet-compressed data are available, but they are not encountered in such a research study that uses these compressed data for classification. Compression takes some time, but it decreases the data that are processed. Thus, it is faster to use compressed data, especially in big databases. If we think about wireless transmission of the signal, it is even better to compress and send. The data were preprocessed, similar to most of the research for improved classification [1–

27]. Denoising, trend elimination, baseline correction, and normalization were found to be necessary for better performance. The usage of wavelets for feature extraction is a known approach. However, as an extra step, we reduced the number of features by deciding on coefficients that have the most significant effect on classification and usage of PCA [38]. Even though PCA is a method used by researchers, checking each coefficient one by one and using PCA at the same time is not common. After deciding on coefficients and adding the standard deviation and variance of the feature set, PCA was applied to finalize our feature set. The dimension of the feature set was six. Our feature set is smaller than those of other related works. It should be noted that two small windows are used when calculating wavelet coefficients. Even though the size of the window differs in each study, most research uses one window [1–27]. Usage of two small windows reduces the number of features, which results in faster classification.

Finally, we use bagged trees for classification to compare the performance of 12-lead ECG recordings. Basic classification methods were compared to determine if a better method was available for our research (as seen in Table 2). The wavelet coefficients were also classified using spiking neural networks, and the accuracy of normalized and unnormalized systems was obtained.

Generally, neural networks are preferred for classification [1–3,6,11,12,19,25,27]. The SVM [4,5,8,14,16,18, 21,22,24] is another popular technique for classification of arrhythmia. Random forest [9], linear classifier [10], morphology consistency evaluation [13], cluster and centroid identification [15], linear discriminant analysis [17], threshold based classifier [20], KNN [7,23], and naive Bayes [26] are the other classification methods that are used in related works. The results obtained from 12-lead ECG recordings using bagged trees classification showed that our performance was as successful as others and more stable. The results obtained from 12-lead ECG recordings using bagged trees classification can be seen in Table 4.

Table 4. The results obtained from 12-lead recordings using bagged trees classification.

Bagged trees	Training	Test		
	Accuracy	Accuracy	Sensitivity	Specificity
Lead 1	98.9	97.8	87	99
Lead 2	99.0	97.9	88	99
Lead 3	97.9	96.6	81	99
Lead 4	99.1	98.7	93	99
Lead 5	98.2	97.1	82	99
Lead 6	98.7	98.1	88	99
Lead 7	98.8	98.1	90	99
Lead 8	98.9	98.1	89	99
Lead 9	98.5	97.9	87	99
Lead 10	98.3	97.7	87	99
Lead 11	98.3	97.8	88	99
Lead 12	98.5	98.0	89	99

10. Conclusion and future works

This research proposes a fast, simple, and accurate arrhythmia detection system based on wavelet parameters. The bagged tree method was chosen for classification. Simulation in MATLAB R2016a showed that our system achieves 98.6% accuracy for training and 97.8% accuracy, 87.4% sensitivity, and 99.7% specificity for testing (average values) with wavelet-compressed data from the St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database. Our system was as successful as others and more stable. Also, our method is found to be fast enough for real-time applications because of a significantly reduced number of features. This reduction was achieved by using two small windows when calculating wavelet coefficients and finding the coefficients that had the most significant effect on classification. We used PCA for further reduction of the number of features. In future studies, new and more extensive databases are planned to be used with different features combined with wavelet coefficients.

Acknowledgment

This research was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant Number 114E452.

References

[1] Dokur Z, Olmez T, Yazgan E. ECG waveform classification using the neural network and wavelet transform. In: Proceedings of the First Joint BMES/EMBS Conference. Vol. 1; 13–16 October 1999; Atlanta, GA, USA. New York, NY, USA: IEEE. p. 273.

[2] Prasad GK, Sahambi JS. Classification of ECG arrhythmias using multi-resolution analysis and neural networks. In: Conference on Convergent Technologies for the Asia Pacific Region; 15–17 October 2003; Bangalore, India. New York, NY, USA: IEEE. pp. 227-231.

[3] Shyu LY, Wu YH, Hu W. Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG. IEEE T Bio-Med Eng 2004; 51: 1269-1273.

[4] Zhao Q, Zhang L. ECG feature extraction and classification using wavelet transform and support vector machines. In: International Conference on Neural Networks and Brain; 13–15 October 2005; Beijing, China. New York, NY, USA: IEEE. pp. 1089-1092.

[5] Jiang X, Zhang L, Zhao Q, Albayrak, S. ECG arrhythmias recognition system based on independent component analysis feature extraction. In: 2006 IEEE Region 10 Conference; 14–17 November 2006; Hong Kong. New York, NY, USA: IEEE. pp. 1-4.

[6] Yu SN, Chen YH. Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. Pattern Recogn Lett 2007; 28: 1142-1150.

[7] İşler Y, Kuntalp M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. Comput Biol Med 2007; 37: 1502-1510.

[8] Zellmer E, Shang F, Zhang H. Highly accurate ECG beat classification based on continuous wavelet transformation and multiple support vector machine classifiers. In: 2nd International Conference on Biomedical Engineering and Informatics; 17–19 October 2009; Tianjin, China. New York, NY, USA: IEEE. pp. 1-5.

[9] Emanet N. ECG beat classification by using discrete wavelet transform and Random Forest algorithm. In: Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control; 2–4 September 2009; Famagusta, North Cyprus. New York, NY, USA: IEEE. pp. 1-4.

[10] Llamedo M, Martínez JP. Heartbeat classification using feature selection driven by database generalization criteria. IEEE T Bio-Med Eng 2011; 58: 616-625.

[11] Rai HM, Trivedi A, Shukla S, Dubey V. ECG arrhythmia classification using Daubechies wavelet and radial basis function neural network. In: 2012 Nirma University International Conference on Engineering; 6–8 December 2012; Ahmedabad, India. New York, NY, USA: IEEE. pp. 1-6.

[12] Rai HM, Trivedi A. ECG signal classification using wavelet transform and back propagation neural network. In: 5th International Conference on Computers and Devices for Communication; 17–19 December 2012; Kolkata, India. New York, NY, USA: IEEE. pp. 1-4.

[13] Li Y, Bisera J, Weil MH, Tang W. An algorithm used for ventricular fibrillation detection without interrupting chest compression. IEEE T Bio-Med Eng 2012; 59: 78-86.

[14] Ye C, Kumar BV, Coimbra MT. Heartbeat classification using morphological and dynamic features of ECG signals. IEEE T Bio-Med Eng 2012; 59: 2930-2941.

[15] Llamedo M, Martínez JP. An automatic patient-adapted ECG heartbeat classifier allowing expert assistance. IEEE T Bio-Med Eng 2012; 59: 2312-2320.

[16] Faziludeen S, Sabiq PV. ECG beat classification using wavelets and SVM. In: 2013 IEEE Conference on Information & Communication Technologies; 11–12 April 2013; Thuckalay, India. New York, NY, USA: IEEE. pp. 815-818.

[17] Chen T, Mazomenos EB, Maharatna K, Dasmahapatra S, Niranjan M. Design of a low-power on-body ECG classifier for remote cardiovascular monitoring systems. IEEE J Em Sel Top C 2013; 3: 75-85.

[18] Dima SM, Panagiotou C, Mazomenos EB, Rosengarten JA, Maharatna K, Gialelis JV, Curzen N, Morgan J. On the detection of myocardial scar based on ECG/VCG analysis. IEEE T Bio-Med Eng 2013; 60: 3399-3409.

[19] Saminu S, Özkurt N, Karaye I. A. Wavelet feature extraction for ECG beat classification. In: IEEE 6th International Conference on Adaptive Science & Technology; 29–31 October 2014; Ota, Nigeria. New York, NY, USA: IEEE. pp. 1-6.

[20] Banerjee S, Mitra M. Application of cross wavelet transform for ECG pattern analysis and classification. IEEE T Instrum Meas 2014; 63: 326-333.

[21] Sahoo SK, Subudhi A, Kanungo B, Sabut SK. Feature extraction of ECG signal based on wavelet transform for arrhythmia detection. In: 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization; 24–25 January 2015; Visakhapatnam, India. New York, NY, USA: IEEE. pp. 1-5.

[22] Jacob N, Joseph, LA. Classification of ECG beats using cross wavelet transform and support vector machines. In: IEEE Recent Advances in Intelligent Computational Systems; 10–12 December 2015; Trivandrum, India. New York, NY, USA: IEEE. pp. 191-194.

[23] Saini R, Bindal N, Bansal P. Classification of heart diseases from ECG signals using wavelet transform and kNN classifier. In: 2015 International Conference on Computing, Communication & Automation; 15–16 May 2015; Noida, India. New York, NY, USA: IEEE. pp. 1208-1215.

[24] Sharma LN, Tripathy RK, Dandapat S. Multiscale energy and eigenspace approach to detection and localization of myocardial infarction. IEEE T Bio-Med Eng 2015; 62: 1827-1837.

[25] Dewangan NK, Shukla SP. ECG arrhythmia classification using discrete wavelet transform and artificial neural network. In: IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology; 20–21 May 2016; Bangalore, India. New York, NY, USA: IEEE. pp. 1892-1896.

[26] Sayilgan E, Cura ÖK, İşler Y. Use of clustering algorithms and extreme learning machine in determining arrhythmia types. In: 25th Signal Processing and Communications Applications Conference; 15–18 May 2017; Antalya, Turkey. New York, NY, USA: IEEE. pp. 1-4.

[27] Rad AB, Eftestol T, Engan K, Irusta U, Kvaloy JT, Kramer-Johansen J, Wik L, Katsaggelos AK. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. IEEE T Bio-Med Eng 2017; 64: 2411-2418.

[28] Graps A. An introduction to wavelets. IEEE Comput Sci Eng 1995, 2: 50-61.

[29] David VK, Rajasekaran S. Pattern Recognition Using Neural and Functional Networks. Berlin, Germany: Springer, 2009.

[30] Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. Cambridge, MA, USA: Academic Press, 2008.

[31] Akansu AN, Haddad RA. Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets. 2nd ed. Cambridge, MA, USA: Academic Press, 2001.

[32] Bronzino JD. Medical Devices and Systems. 3rd ed. Boca Raton, FL, USA: CRC Press, 2006.

[33] Jolliffe IT. Principal Component Analysis. New York, NY, USA: Springer, 1986.

[34] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000; 101: 215-220.

[35] Breiman L. Bagging predictors. Mach Learn 1996; 24: 123-140.

[36] Diehl PU, Neil D, Binas J, Cook M, Liu SC, Pfeiffer M. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In: 2015 International Joint Conference on Neural Networks; 12–17 July 2015; Killarney, Ireland. New York, NY, USA: IEEE. pp. 1-8.

[37] Jha CK, Kolekar MH. Performance analysis of ECG data compression using wavelet based hybrid transform method. In: 2015 International Conference on Microwave, Optical and Communication Engineering; 18–20 December 2015; Bhubaneswar, India. New York, NY, USA: IEEE. pp. 138-141.

[38] Kim J, Shin H, Lee Y, Lee M. Algorithm for classifying arrhythmia using extreme learning machine and principal component analysis. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 22–26 August 2007; Lyon, France. New York, NY, USA: IEEE. pp. 3257-3260.