Anatolian Journal of Computer Sciences

Volume:5 No:1 2020 © Anatolian Science

pp:1-13 ISSN:2548-1304

Heart Sound Classification for Murmur Abnormality Detection Using an Ensemble Approach Based on Traditional Classifiers and Feature Sets

Ali Fatih Gündüz[1], Ali Karcı[2]*

*[*1]Akçadağ Voc. Highschool, Malatya Turgut Ozal Uni., Malatya, Turkey (fatih.gunduz@ozal.edu.tr)*

[2]Department of Computer Eng., Inonu University, Malatya, Turkey, (ali.karci@inonu.edu.tr)

Received Date: Jan. 7, 2020.
Acceptance Date: Mar. 21, 2020.
Published Date : Jun. 1, 2020.

Abstract— Phonocardiography (PCG) is a method based on examination of mechanical sounds coming from heart during its regular contraction/relaxation activities such as opening and closing of the valves and blood turbulence towards vessels and heart chambers. Today there are high technology tools to record those sounds in electronic environment and enable us to analyze them in detail. The constraints such as human's limited audible range, environment noise and inexperience of physicians can be overcome by the use of those tools and development of state-of-art signal processing and machine learning methods. In this study we examined heart sounds and classified them as normal or abnormal focusing on the efficiency of ensemble classifiers. Features of heart sounds are extracted by using Discrete Wavelet Transform (DWT), Mel-Frequency Cepstral Coefficients (MFCC) and time-domain morphological characteristics of the signals. As a novel contribution, Karcı entropy is derived from DWT of PCG signals and used for the first time as feature in heart sound classification. K-Nearest Neighbor (kNN), Support Vector Machine (SVM), Multilayer Perceptron (MLP) classifiers and their ensembles are used as classifiers. Then the ensemble classifiers' predictions based on distinct feature vectors are combined and an ensemble classifier built from team of ensemble classifiers. Classification performances of singular classifiers, single level ensemble classifiers and final ensemble classifier are compared and better results are obtained by the proposed method.

Keywords : *Karcı entropy, signal processing, data mining, classification, heart sound*

1.Introduction and Related Work

Heart and blood vessels related disorders are called as cardiovascular diseases (CVDs) which include arrhythmia, heart attacks and strokes. According to World Health Organization (WHO), CVDs are the most common causes of death globally and most people die due to CVDs annually than any other causes [1]. Even though exact causes of CVDs are not conceded clearly by authorities, there are some accepted risk factors such as hypertension, diabetes, obesity, smoking and family history of having CVD. People having the risk factors need to be more careful and concerned about their well-being. Early identification of heart diseases with high accuracy provides vital information and can be extremely important for survival of those people [2].

Doppler-Echocardiography, Magnetic Resonance Imaging (MRI), Electrocardiography (ECG) and Phonocardiogram (PCG) are effective tools for measuring dysfunctions of heart. Those methods provide accurate results although they require expensive equipment which can be used by expert physicians. On the other hand, auscultation is a classical method which is defined as listening to the cardiac sounds with a stethoscope. Auscultation is simple, practical, noninvasive and cost-effective therefore it is more suitable for home and primary health care. A lot of cardiac diseases are identified firstly by finding presences of abnormalities in the heart sounds [3] however, differentiating pathological symptoms from normal heart sounds requires knowledge and experience. Moreover, it can be a subjective task even among clinicians [4, 5].

In order to obtain more successful and objective evaluation of heart sounds, physicians make use of computer-based automatic systems for heart sound diagnosis in clinics. Thanks to development of technology, today there are portable heart sound evaluation devices corroborated with digital signal analysis methods. Many state-of-art methods [6-10] have been proposed to meet this need. Those tools can be used to educate and train cardiology students and inexperienced physicians. Also, they can be very helpful in rural areas and in homecare where it is not easy to reach expert consultation [2].

During cardiac cycle, a healthy heart generates a regular pair of sounds called S1 and S2 at each pulsation due to rhythmic opening and closing of heart valves [11]. S1 and S2 are also known as fundamental heart sounds and are described as lub-dup in layman's terms. Variation in their durations and intensities are regarded as important implications of cardiac disorders [8]. The third heart sound S3, also known as ventricular gallop, follows S2 and in most adults it is soft and subdued. However, it can be loud enough to be heard among children, pregnant females and well-trained athletes without indicating any cardiac abnormality. On the other hand, another heart sound S4, which occurs just before S1, must be inaudible in healthy subjects. Moreover, in some pathological cases those sounds are accompanied by a noise which called as heart murmurs. Having heart murmurs is not a disease but obviously indicates an abnormal turbulence of blood flow either inside heart or in vessels. Sometimes, its presence may be normal without pointing out any cardiac pathology in which case it is called as "innocent murmur" [3,12]. Detecting and classifying murmurs during auscultation is an important task for a physician.

Murmurs can be caused either from inborn structural defects of the heart or from physiologic conditions which take place subsequently. Structural defects related with heart valves include stenosis (heart valve flaps restrict the blood flow by fusing together), regurgitation (flaps do not close properly allowing back flow of blood) and atresia (valves do not develop congenitally preventing it from opening) [13]. On the other hand, pregnancy, fever, anemia or post-operation effects can be counted as external causes of heart murmurs occurring later.

There are various types of murmurs and they can take place in different locations of a pulse depending on their kind. Sometimes being sub-audible notwithstanding, they can be detected and seen in PCG recordings [13]. As a gold standard for grading heart murmur intensity, murmurs are graded on a 6-point level scale (Levine scale) by clinicians [14]. Separating heart sounds from murmurs can be challenging since they involve low frequency components which are hardly audible [15]. Especially presence of environmental noise, breathing sounds, rustling of the microphone due to movement make it harder for naked human ear. Therefore, there is a need for developing human-machine interfaces to help cardiologists understand and interpret PCG outputs more accurately and easily during the diagnosis [16]. Actually, automated classification of heart sounds has been studied for decades however, achieving more accurate classification results is still attracting researchers [10].

Munia et. al. [17] used Support Vector Machine (SVM) for classifying heart sounds into normal and abnormal classes by comparing five different kernel functions. Koçyiğit [18] proposed a classification method based on obtaining sub-bands of heart sounds by using DWT. Features are extracted from those sub-bands and then dimensions of feature vectors are reduced by using principal component analysis on them. Those features are given to Naïve Bayes and SVM classifiers to obtain fourteen different heart sound classes. In another study, Uzunhisarcıklı [19] examined the blood flow velocity variation due to mitral valve stenosis by performing an analysis on Doppler ultrasound signals and extracted nonlinear features from the heart sound for this purpose.

Potes et. al. [7] proposed an ensemble classifier method combined from AdaBoost and Convolutional Neural Network (CNN) which are trained with time-domain and frequency-domain features. Zabihi et. al. [8] used an ensemble of 20 feedforward Neural Networks which are built from two hidden layers containing 25 neurons. They used Linear Predictive Coefficients (LPC), entropy, Mel-Frequency Cepstral Coefficients (MFCC), Discrete Wavelet Transform (DWT) and Power Spectral Density (PSD) to extract features. In another study, Kay and Agarwal [9] proposed a Neural Network based method using Continuous Wavelet Transform (CWT) and MFCC for feature extraction. In feature vectors, they also used time-domain features obtained from the segmented signal such as mean, standard deviation and inter-segment ratios.

Each regular pulse of heart generates a cardiac cycle which is created from distinct temporal regions. The time interval from completion of S1 to initialization of S2 is called as systole and S2 to S1 period is known as diastole [20]. Partitioning the signal into repeating individual segments is an important step of automated heart sound analysis which can be achieved by using threshold-based change point detection methods or probabilistic models [21]. Moreover, fundamental heart sound segmentation can be done by using extra reference signals such as ECG when it is recorded in parallel to PCG. Indicating cardiac cycles by R peaks of QRS complexes, ECG makes it easy to describe temporal segments of the PCG record. Segmentation can also be achieved by using different envelope extraction methods like Hilbert-Huang transform [22], Shannon energy [6] and autocorrelation [23]. Without using any reference signal, Springer et. al. [21] proposed a hidden semi-Markov model extended with logistic regression for determining four states of heart cycle: S1, S2, systole and diastole. However, there are successful heart sound analysis approaches [8, 24, 25] that do not involve segmentation at all.

Another necessary step of heart sound classification is feature extraction. Frequency, time and time-frequency domain features are main characteristics of heart sounds [26]. Heart sounds are observed in a specific frequency level. Frequency components are in the range of 10-140 Hz for S1, 10-200 Hz for S2 and 20-70 Hz for S3 and S4 [26]. According to another study [27] fundamental heart sounds should be expected at the range of 20-200 Hz but this frequency spectrum can be 15-700 Hz in pathological cases. Murmurs, on the other hand, can have diverse frequency components up to 600 Hz [26, 28] but this level can increase up to 1000 Hz occasionally [28]. Time domain features include durations of segments and their ratio to overall signal, energy measures, zero-crossing rate, minimum, maximum, mean, variance, skewness, kurtosis and Lempel-Ziv complexity (LZC) [15, 29]. Lastly, time-frequency domain features are computed from windowed segments of the signal. By this approach it is possible to observe frequency features at specific time intervals. Those features are used to detect how frequency content of the signal changes by time. Short-Time Fourier Transform (STFT), LPC and MFCC, DWT and CWT are methods used to produce time-frequency domain features [8, 30]. Among those alternatives, wavelet transforms get upper hand because of their ability for analyzing non-stationary signals and PCG signals are non-stationary in nature like many other bio-signals [16, 17].

The aim of this study is to propose a three-phase model for automatic heart sound classification to detect heart murmurs. MFCC, time domain morphological characteristics and entropies calculated from DWT details and approximation coefficients of heart sounds are used to form three different feature vectors in this study. Instead of merging them together, they are examined individually. For this purpose, classification algorithms K-Nearest Neighbor (kNN), Multilayer Perceptron (MLP) and SVM are used at first. During the first phase, those classifiers are run with mentioned feature vectors separately. In the second phase, those classifiers are combined together and their ensemble is run with those feature vectors one by one. Then in the third phase, ensemble classifier's prediction decisions for individual feature vectors are combined constructing a single classifier from the team of the ensemble classifiers. In the next section, the database and the proposed method are explained. Experiment results are given in the Section 3. Finally, Section 4 details the conclusions of the study.

2. Database and Proposed Method

In 2016, PhysioNet/Computing in Cardiology (CinC) announced a competition challenge for encouraging development of heart sound classification algorithms and provided an online database of PCG recordings [26, 31]. Researchers used the shared database to train and test their algorithms however submissions were run by using a hidden database and ranked according to their performances by competition organizers. The public part of this challenge database is used in this study. It totally contains 3240 wav files of which 2575 recordings are normal and 665 recordings are abnormal. Details about diseases are not given and all sounds are categorized either as normal, abnormal or uncertain. The heart sounds are recorded with 2 kHz sampling frequency.

The proposed method is composed of signal preprocessing, feature extraction and classification steps which are detailed in the following subsections. A total of 38 features extracted from each PCG record which are shown in the Table 1. Features based upon morphological properties, MFCC and DWT detail and approximation coefficients are extracted and analyzed independently. As a novel contribution, Karcı entropy is derived from DWT of PCG signals and used for the first time as feature in heart sound classification.

The features were not combined into a single feature vector. Instead they were used to form three distinct feature vectors namely TD (time domain), MFCC and DWT based features. The prediction results of ensemble classifiers were combined further by majority voting rule to obtain a final ensemble classifier.

The main contributions of our work include investigation of the effect of different features on the classification performance. Moreover, Karcı entropy [32] is used for the first time for heart sound analysis. Those features are used to train kNN, SVM and MLP classifiers. Another contribution of our study is three-level classification approach. Those classifiers are compared among themselves in the first level and their ensembles are built. In the second level ensemble classifiers are supplied with mentioned feature vectors and classification performances are compared. Then in the third level, final ensemble classifier is built from the team of ensemble classifiers and overall performances are compared. The general overview of the system is given in Figure 1. As it can be seen in the figure, the proposed method starts with preprocessing, and then extracts three different feature vectors from PCG signals. Those feature vectors are classified by kNN, SVM, MLP and their ensembles.

Table 1. Feature sets and their elements

Feature Set	Features
Time Domain Based Features	Mean, Std, LZC, Energy, Zero-cross rate and Delta (6)
MFCC Based Features	Average (1-13), Skewness (14-26)
DWT Based Features	5 level details (1-5), approximation (6)

2.1. Preprocessing

The heart sounds are recorded from distinct locations by using different acquisition techniques. In order to get rid of the variations in the amplitude, the data should be normalized. After calculating and storing the mean and standard deviation of the signal, it is normalized by subtracting its mean and dividing by its standard deviation (z-score normalization). For all features except delta, segmentation is not used and the features are extracted from whole signal during feature engineering. For delta feature calculation, the signal is converted to 2-dimensional input model in which frame length is kept as 1 second. The last frames of the signals having duration less than 1 second are disposed. The remaining frames formed the 2-dimensional representation of the preprocessed signal.

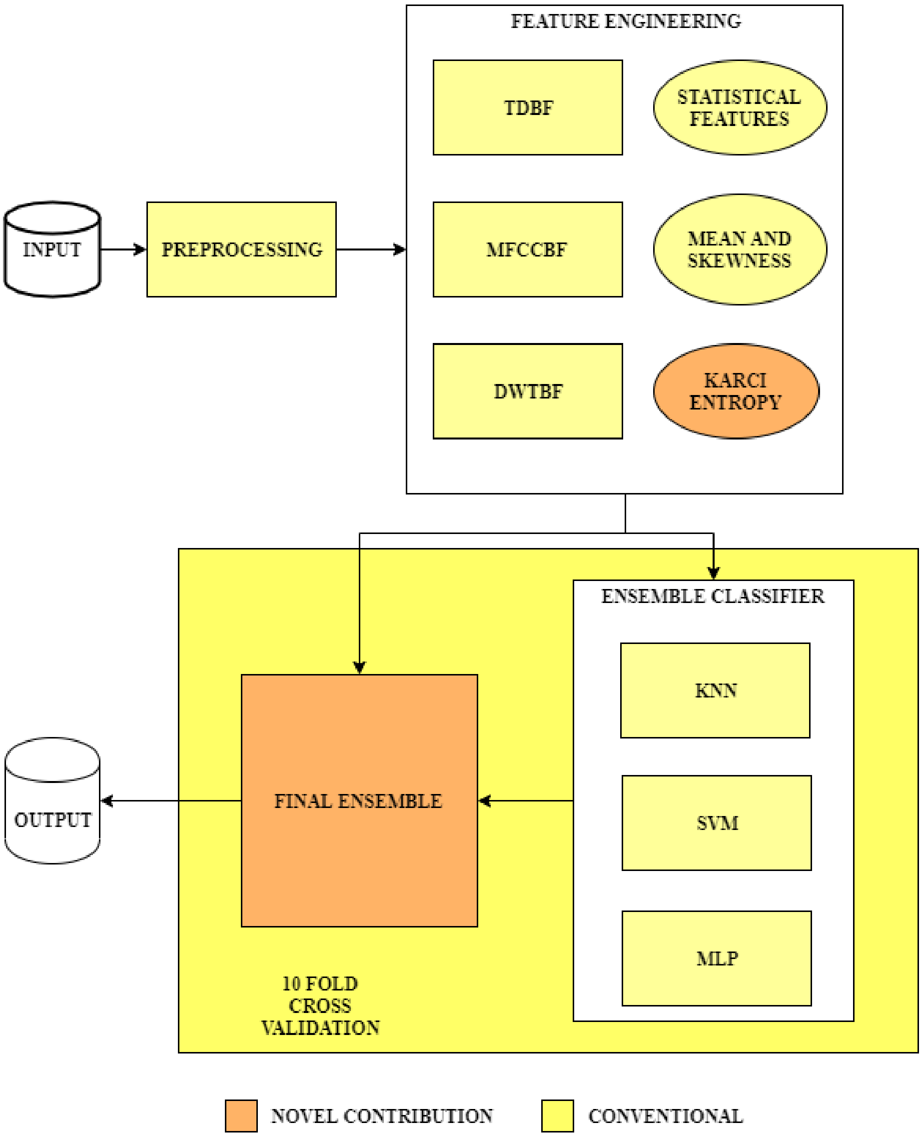

Figure 1. General overview of the system and the proposed method

2.2. Feature Engineering

2.2.1. Time Domain Based Features (TDBF)

The mean and standard deviation obtained from raw data are added to feature vector. Then Lempel-Ziv complexity (LZC) is calculated from the signal. LZC is a non-linear feature estimating the repeating patterns in a signal. Non-stationary signals get higher LZC scores while periodic signals have a low score [29].

To calculate LZC, a new sequence (x) denoted as x(1), x(2), …, x(n) should be created from the original signal (s) where n is the length of the signal. An element of x at active index i, is set to 1 if s(i+1) > s(i) and in other cases x(i) is set to 0. By following this way, the x sequence is used to count increase points of the s. The final LZC is obtained according to Equation 1. Shannon energy [6] is another property added to feature vector. Shannon energy of a normalized signal (s_{norm}) is estimated according to Equation 2. Then the number of times the signal changes its sign from positive to negative or from negative to positive is counted. The sum of the changes is divided by the total length to calculate the zero-cross rate (ZCR). According to [33], low ZCR is expected for normal signal while high ZCR indicates abnormality. ZCR features can be estimated according to Equation 3. Delta score (Δ) of a signal is calculated by averaging the sum of differences between maximum and minimum of each frame of the signal, which is given in Equation 5.

$$\text{LZC (s)} = \frac{\sum x(i) * \log(n)}{n} \tag{1}$$

$$\text{E}(s_{norm}) = \frac{-1}{n} * \sum s^2_{norm}(i) * \log(s^2_{norm}(i)) \tag{2}$$

$$\text{ZCR(s)} = \frac{1}{2n} * \sum \left|\text{sgn}(\text{s(i)}) - \text{sgn}(\text{s(i} - 1))\right| \tag{3}$$

$$\text{sgn(x(n))} = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \tag{4}$$

$$\Delta(s) = \frac{1}{Frames\_Count} * \sum_F^{All\ Frames}[\max(F) - \min(F)] \tag{5}$$

2.2.2. Mel-Frequency Cepstral Coefficients Based Features (MFCCBF)

Speech recognition studies widely use MFCC [34] and they are proven to be distinctive features for heart sound classification as well [9]. Human ear pays different levels of regard to different frequencies. MFCC is designed in a way to represent the sound signals as a human would represent them considering the frequency levels. In order to obtain the coefficients, the signal is divided into overlapping frames which are exposed to Hamming window minimizing the discontinuity. By applying Fourier Transform, they are transferred into frequency domain then they are passed from Mel-spaced filter bank. Finally, discrete cosine transform is applied on the result to convert it back to time domain. Depending upon the sampling frequency, predefined number of coefficients are obtained from each segment [35]. In this study, MFCCs of signals are calculated with 13 coefficient parameters. Overlap length is selected as 10 ms and frame length is taken as 20 ms. From those sets of 13 coefficients, average values and skewness values are calculated. A total of 26 features are used to create MFCC based feature vectors.

2.2.3. Discrete Wavelet Transform Based Features (DWTBF)

Daubechies 4 (db4) wavelet function is applied on signals during discrete wavelet transform. Detail coefficients of level 1 to 5 and approximation coefficients are saved. Then those coefficients' probability density functions (p_i) are estimated. As a novel contribution, Karcı entropy is derived from those probability density functions and used for the first time as feature in heart sound classification.

Karcı entropy which is given in Equation 6, is superior to widely used Shannon entropy since it is a larger set that contains Shannon Entropy. As it can be seen in Equation 6, there is an (α) parameter, which can be selected based on a fuzzy approach [36]. When its value is set to 1, it works exactly like Shannon entropy. This parameter is used for fine-tuning the entropy calculation and the performance of the overall system can show different behavior due to its value.

In this study, we experimentally searched for best α parameter in the range of [0.5-5.5] by 0.01 incrementations. As a result of those experiments 1.14 is selected as α value to use in the calculation of Karcı entropy. At the end, 6 entropy-based attributes are extracted from details 1-5 and approximation of discrete wavelet transform and then added to the feature vector.

$$Karcı\ Entropy = \sum_i^n |(-p_i)^\alpha * \ln(p_i)| \tag{6}$$

2.3. Classification

Classification phase of the study is realized by using WEKA workbench [37]. For this purpose, famous classifiers kNN, SVM and MLP are supplied with the feature vectors recorded in WEKA arff format. For train and test phases of classification, 10-fold cross validation strategy is followed. Each feature vector is separately supplied to kNN, SVM and MLP classifiers and their performances are compared. Then, instead of concatenating all those features and training the classifiers with it, an ensemble approach is proposed. An ensemble classifier can be constructed from different classifiers (or same classification algorithm with distinct parameters) using same input features or can also be built from classifiers each using its own input feature set [38].

In this study, the both approaches are combined by developing a three-level classification phase. Firstly, kNN, SVM and MLP is run with same input feature vectors and their classification performances are recorded. Then an ensemble of those classifiers is built and run three times with three different feature vectors. Finally, ensemble classifier’s predictions about particular objects are improved by combining three classification decisions into one by using majority vote approach.

WEKA has an effective and simple ensemble implementation, Vote[1], capable of combining two or more classifiers. When classifiers produce models assigning probabilities to label each object, it is possible to get an overall prediction from those probabilities by applying different combination rules [39, 40]. Those rules can be maximum, minimum, median, average probability, product of probabilities and majority vote. In this study, majority vote combination rule is preferred and applied.

2.3.1. K-Nearest Neighbor (kNN)

kNN is an intuitive, widely used, instance based and non-parametric supervised learning algorithm. It is advantageous in many cases because it does not rely on an assumption about data’s belonging to a particular family of distribution. The algorithm classifies objects by checking the most common class among its k neighbors where k is a positive and generally odd integer [41]. Calculating distances between each pair of objects in order to determine neighbors, classification time is directly related with the size of the dataset.

The value of k plays a key role in the performance of the classifier and its best value may change in different data sets. To determine the k value, we conducted a series of experiments and the best performance is obtained with k=9. Therefore, in this study all kNN classifications are done with k parameter as 9.

2.3.2. Support Vector Machine (SVM)

SVM is a useful tool used in pattern recognition and regression analysis. In WEKA environment, there exists a version of it named as SMO which implements John Platt's [42] Sequential Minimal Optimization algorithm for training a support vector classifier. In WEKA, there is another SVM implementation alternatively called as LibSVM however SMO comes built-in with WEKA framework while its alternative requires additional setup. In SMO, the training phase of SVM is iteratively reduced to smaller chunks producing an optimization problem solvable analytically [43]. Then SVM builds a model which assigns objects to classes aiming to keep categories as much divided as possible [44]. Multi-class problems are solved using one-against-one approach in this implementation [45].

Basically, being a linear classifier, SVM works with linearly separable features however feature vectors are not always linearly separable. This problem is overcome with kernel trick. SVM gives more successful results than other classifiers in many cases so long as proper kernels are selected [46]. By using kernel trick, the original input space is mapped into high dimensional space in which features are linearly separable. Polynomial, sigmoid, radial basis function (RBF), linear and Pearson vii function-based universal kernel (PUK) are popular kernel types. In this study, best performance is obtained with PUK and it is used as kernel function.

[1] http://weka.sourceforge.net/doc.dev/weka/classifiers/meta/Vote.html

2.3.3. Multilayer Perceptron (MLP)

MLP is a class of feed-forward artificial neural network and it is useful especially for distinguishing data which are not linearly separable [35]. It consists of input, output and one or more hidden layers which are interconnected. Those layers are set with nodes called as neurons. Those neurons are connected with each other through certain weights and form a network topology. When a neuron is stimulated with an input signal bigger than threshold, it responds back by exciting its connected-neighbors by using an activation function [46]. As training occurs, MLP learning takes place by changing the connection weights. The amount of the step-size weight change is determined by learning rate. In this study, learning rate is used as 0.3 and 3 hidden layers with 5, 10, 15 neurons in each of them are placed in the network.

2.4. Performance Measurement

Although the Physionet challengers defined some metrics based on defined weights for assessing rank of submissions [10], performance results are not computed and compared according to that scheme in this study. Therefore, our proposed method is not directly comparable with the competitors'. The performance results of classifiers, their ensembles and final combination of ensembles are shown and compared in the next section.

True positive (TP) and true negative (TN) metrics are the number of correct positive and negative predictions respectively. False positive (FP) is the number of misclassified positive predictions and false negative (FN) represents the number of negative predictions which are actually positive. Those values are used to build confusion matrix and many other performance evaluators are extracted from them. The ratio of total number of correct predictions to the total number of input samples is called accuracy. Its formula is given in Equation 7. To get a positive assessment, accuracy is the first parameter to check however only using accuracy may be misleading [47]. The imbalanced data sets having significant disparity between distinct classes lead to deceptive conclusions about classifiers when only accuracy results are considered.

Precision and recall are two metrics used to overcome class-imbalance problem. Precision, which is also called positive predictive value, is the ratio of number of correct predictions to the number of predictions. A low precision indicates high false positives with respect to true positives. On the other hand, recall is the ratio of number of correct predictions to the total number of the actual members of the class in the question. The effect of high false negative presence results in low recall scores.

F1 score, which is also called F-measure, is used to balance between precision and recall [48] and it is calculated by taking harmonic mean of them. Its formula is given in Equation 10. Combining precision and recall into a single measure of the performance, F1 score is commonly used but those three metrics are criticized because of the fact that true negatives are ignored [49]. Precision, recall and F1 score are calculated for both normal and abnormal classes in order to pay attention to the negatives and to obtain a better estimation of overall performance. Weighted averages of those two measurements are also calculated.

$$\text{Accuracy } = \frac{TP+TN}{TP+FP+TN+FN} \tag{7}$$

$$\text{Precision } = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall } = \frac{TP}{TP+FN} \tag{9}$$

$$\text{F1 Score} = \frac{2*Precision*Recall}{Precision+Recall} \tag{10}$$

3. Experiment Results

Using heart sound database of PhysioNet/CinC 2016 challenge, a comparative study between kNN, SVM, MLP classifiers and their ensembles is carried out. Three kinds of feature sets are investigated and ten-fold cross validation is applied. The proposed final ensemble classifier is obtained by combining predictions of the ensemble classifiers for all feature vectors.

Table 2 shows experiment results of kNN, SVM and MLP classifiers with three feature sets DWTBF, MFCCBF and TDBF. Best accuracy among them belongs to SVM using MFCCBF. Best normal class precision is obtained by MLP using MFCCBF while best abnormal class precision is reached by SVM using DWTBF. But best weighted precision score is 0.90 reached by kNN using MFCCBF and SVM using MFCCBF. Best normal class recall score is obtained by SVM using DWTBF however worst abnormal class recall is seen in this experiment as well. Best abnormal class recall is obtained by MLP using MFCCBF. In two experiments, kNN using MFCCBF and SVM using MFCCBF, 0.90 is reached for weighted recall score outperforming the rest. Finally, best normal class F1 score is obtained by kNN using MFCCBF and SVM using MFCCBF. Best abnormal class F1 score is reached in three different experiments: kNN using MFCCBF, SVM using MFCCBF and MLP using MFCCBF. Again, best weighted F1 score is obtained in two of those experiments: kNN using MFCCBF and SVM using MFCCBF. Considering the overall experiment results, the best feature set among the feature vectors is appeared to be MFCCBF and best classifier algorithm is seen as SVM.

Performance results of ensemble of kNN, SVM and MLP classifiers with different feature vectors and proposed method are given in Table 3. According to those results, most of the criteria (accuracy, abnormal class precision, weighted average of precisions, normal class recall, weighted average of recalls, normal class F1 score, abnormal class F1 score and weighted average of F1 scores) favor the proposed method. Only best normal class precision and abnormal class recall scores are obtained by single level ensemble classifier using MFCCBF. The single level ensemble classifiers using the other feature vectors (TDBF and DWTBF) are left behind the one using MFCCBF. When we compare the performance results between the best ensemble with the best singular classifier, the overall performance is slightly improved. Building an ensemble from the team of ensemble classifiers improves the overall performance further.

Table 2. Classifiers' accuracy percent (Acc.), normal class precision (Pre.N.), abnormal class precision (Pre.A.), weighted precision (Pre.W.), normal class recall (Rec.N.), abnormal class recall (Rec.A), weighted recall (Rec.W.), normal class F1 score (F.N.), abnormal class F1 score (F.A.) and weighted F1 score (F.W.). Best result of each row is in bold.

Method	**kNN**			**SVM**			**MLP**		
Features:	DWTBF	MFCCBF	TDBF	DWTBF	MFCCBF	TDBF	DWTBF	MFCCBF	TDBF
Acc(%)	85.00	90.06	88.21	80.80	**90.12**	87.28	83.40	88.92	88.00
Pre.N.	0.87	0.92	0.91	0.81	0.91	0.91	0.86	**0.93**	0.90
Pre.A.	0.74	0.83	0.75	**0.91**	0.85	0.71	0.65	0.74	0.77
Pre.W.	0.84	**0.90**	0.88	0.83	**0.90**	0.87	0.82	0.89	0.87
Rec.N.	0.96	0.96	0.95	**1.00**	0.97	0.93	0.94	0.93	0.96
Rec.A.	0.42	0.65	0.64	0.07	0.64	0.65	0.42	**0.72**	0.59
Rec.W.	0.85	**0.90**	0.88	0.80	**0.90**	0.87	0.83	0.89	0.88
F.N.	0.91	**0.94**	0.93	0.89	**0.94**	0.92	0.90	0.93	0.93
F.A.	0.53	**0.73**	0.69	0.13	**0.73**	0.68	0.51	**0.73**	0.67
F.W.	0.83	**0.90**	0.88	0.74	**0.90**	0.87	0.82	0.89	0.87

Table 3. Performances of ensemble classifiers with discrete wavelet transform based features (DWTBF), Mel-Frequency Cepstral Coefficients based features (MFCCBF) and time domain based features (TDBF): accuracy percent, normal class precision (Pre.N.), abnormal class precision (Pre.A.), weighted precision (Pre.W.), normal class recall (Rec.N.), abnormal class recall (Rec.A), weighted recall (Rec.W.), normal class F1 score (F.N.), abnormal class F1 score (F.A.) and weighted F1 score (F.W.). Best result of each row is in bold.

	Ensemble Classifier with DWTBF	**Ensemble Classifier with MFCCBF**	**Ensemble Classifier with TDBF**	**Proposed Approach**
Acc.(%)	84.88	90.56	88.46	**90.93**
Pre.N.	0.86	**0.92**	0.91	0.91
Pre.A.	0.77	0.84	0.76	**0.89**
Pre.W.	0.84	0.90	0.88	**0.91**
Rec.N.	0.97	0.97	0.95	**0.98**
Rec.A.	0.37	**0.67**	0.65	0.64
Rec.W.	0.85	0.90	0.89	**0.91**
F.N.	0.91	**0.94**	0.93	**0.94**
F.A.	0.50	**0.74**	0.70	**0.74**
F.W.	0.83	0.90	0.88	**0.91**

4. Conclusions

Interpreting heart sounds and detecting murmurs accurately is a challenging and interesting task. This study presented an original solution for this problem by using public PCG database provided by Physionet/CinC 2016. The proposed method consists of signal preprocessing, feature engineering and classification phases. In the preprocessing phase, the PCG signals were normalized by subtracting their mean and dividing by their standard deviation. Then 38 features were extracted from those signals by using time domain statistical characteristics of the signal, MFCC and DWT detail and approximation coefficients. Those features were grouped into three sets in order to form three distinct feature vectors.

As a novel contribution, Karcı entropy is used for the first time as feature in heart sound classification. Comprising Shannon entropy, its formula is suited for giving more successful results depending on its alpha parameter. When it is 1.0, Shannon entropy works exactly same as Karcı entropy. This parameter is calculated in a fuzzy logic approach and gets significant values depending on the problem set.

The classification phase of the proposed method was based on a three-level approach during which effectiveness of feature vectors were compared. During the first level, well-known algorithms kNN, SVM and MLP were used one by one with all feature vectors. Then in the second level, ensemble of those classifiers was formed and tested. This ensemble classifier was run three times with three distinct feature vectors. Lastly, in the third level, a final ensemble classifier was built from those three trials. Three sets of predictions obtained from ensemble classifier at the second level were combined by using majority vote principle and final classification decisions were brought forward by this approach. This study also targeted to compare the performances of kNN, SVM and MLP. Moreover, distinct feature vectors' effect on the classification performance was compared as well. Overall accuracy, precision, recall and f1 scores increased at each level of the classification and best results were obtained by the final ensemble classifier.

References

[1] WHO.2017 world statistics on cardiovascular disease. URL https://www.who.int/news-room/factsheets/detail/cardiovascular-diseases-(cvds)

[2] Arslan, Ayse, and Oktay Yildiz. "Automated Auscultative Diagnosis System for Evaluation of Phonocardiogram Signals Associated with Heart Murmur Diseases." Gazi University Journal of Science 31.1 (2018): 112-124.

[3] Thompson, W. Reid, et al. "Artificial intelligence-assisted auscultation of heart murmurs: validation by virtual clinical trial." Pediatric cardiology 40.3 (2019): 623-629.

[4] Delgado-Trejos, Edilson, et al. "Digital auscultation analysis for heart murmur detection." Annals of biomedical engineering 37.2 (2009): 337-353.

[5] El-Segaier, Milad, et al. "Computer-based detection and analysis of heart sound and murmur." Annals of Biomedical Engineering 33.7 (2005): 937-942.

[6] Balili CC, Sobrepena M, Naval PC, Classification of heart sounds using discrete and continuous wavelet transform and random forests. In 3rd IAPR Asian Conference on Pattern Recognition, Kuala Lumpur, 2015.

[7] Potes, Cristhian, et al. "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

[8] Zabihi, Morteza, et al. "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

[9] Kay, Edmund, and Anurag Agarwal. "Dropconnected neural network trained with diverse features for classifying heart sounds." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

[10] Clifford, Gari D., et al. "Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

[11] Quiceno-Manrique, A. F., et al. "Selection of dynamic features based on time{frequency representations for heart murmur detection from phonocardiographic signals." Annals of biomedical engineering 38.1 (2010): 118-137.

[12] Dokur, Zümray, and Tamer Olmez. "Feature determination for heart sounds based on divergence analysis." Digital Signal Processing 19.3 (2009): 521-531.

[13] Chakrabarti, Tamal, et al. "Phonocardiogram signal analysis-practices, trends and challenges: A critical review." 2015 International conference and workshop on computing and communication (IEMCON). IEEE, 2015.

[14] Silverman, Mark E., and Charles F. Wooley. "Samuel A. Levine and the history of grading systolic murmurs." The American journal of cardiology 102.8 (2008): 1107-1110.

[15] Bozkurt, Baris, Ioannis Germanakis, and Yannis Stylianou. "A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection." Computers in biology and medicine 100 (2018): 132-143.

[16] Hanbay, Davut. "An expert system based on least square support vector machines for diagnosis of the valvular heart disease." Expert Systems with Applications 36.3 (2009): 4232-4238.

[17] Munia, Tamanna TK, et al. "Heart sound classification from wavelet decomposed signal using morphological and statistical features." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

[18] Koçyiğit, Yücel. "Heart sound signal classification using fast independent component analysis." Turkish Journal of Electrical Engineering & Computer Sciences 24.4 (2016): 2949-2960.

[19] Uzunhisarcikli, Esma. "Nonlinear dynamic analysis of mitral valve doppler signals: surrogate data analysis." Turkish Journal of Electrical Engineering & Computer Sciences 18.2 (2010): 327-337.

[20] Das, Sangita, Saurabh Pal, and Madhuchhanda Mitra. "Supervised model for Cochleagram feature based fundamental heart sound identification." Biomedical Signal Processing and Control 52 (2019): 32-40.

[21] Springer DB, Tarassenko L, Clifford GD. Logistic regression-HSMM-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 2016;63(4):822{832.

[22] Zhang D, He J, Jiang Y, Du M. Analysis and classification of heart sounds with mechanical prosthetic heart valves based on Hilbert-Huang transform. International Journal of Cardiology2011;151(1):126{127.

[23] Kao WC, Wei CC. Automatic phonocardiograph signal analysis for detecting heart valve disorders. Expert Systems with Applications 2011;38(6):6458{68.

[24] Yuenyong S, Nishihara A, Kongprawechnon W, Tungpimolrut K. A framework for automatic heart sound analysis without segmentation. In BioMedical Engineering OnLine, 2011; 10-13.

[25] Deng SW, Han JQ. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. Future Generation Computer Systems 2016;60:13{21.

[26] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37(9).

[27] Naseri, H., and M. R. Homaeinezhad. "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric." Annals of biomedical engineering 41.2 (2013): 279-292.

[28] Gharehbaghi, Arash, et al. "A novel method for discrimination between innocent and pathological heart murmurs." Medical engineering & physics 37.7 (2015): 674-682.

[29] Cömert, Zafer, and Adnan Fatih Kocamaz. "Evaluation of fetal distress diagnosis during delivery stages based on linear and nonlinear features of fetal heart rate for neural network community." Int. J. Comput. Appl. 156.4 (2016): 26-31.

[30] Ahlstrom, Christer, et al. "Feature extraction for systolic heart murmur classification." Annals of biomedical engineering 34.11 (2006): 1666-1677.

[31] PhysioNet. Classification of normal/abnormal heart sound recordings: the PhysioNet/Computing in Cardiology Challenge 2016[internet].PhysioNet; 2016. Available from https://www.physionet.org/challenge/2016/

[32] Karci, Ali. "Fractional order entropy: New perspectives." Optik 127.20 (2016): 9172-9177.

[33] Homsi, Masun Nabhan, et al. "Automatic heart sound recording classification using a nested set of ensemble algorithms." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

[34] Huang X, Acero A, Hon HW. Spoken Language Processing. A Guide to Theory, Algorithm, and System Development. Prentice Hall, 2001.

[35] Antony, Anett, and R. Gopikakumari. "Speaker identification based on combination of MFCC and UMRT based features." Procedia computer science 143 (2018): 250-257.

[36] Tuğal, Ihsan, and Ali Karcı. "Comparisons of Karcı and Shannon entropies and their effects on centrality of social networks." Physica A: Statistical Mechanics and its Applications 523 (2019): 352-363.

[37] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[38] J. Kittler, M. Hatef, Robert P.W. Duin, J. Matas (1998). On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(3):226-239.

[39] Mondejar-Guerra, V., et al. "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers." Biomedical Signal Processing and Control 47 (2019): 41-48.

[40] Ludmila I. Kuncheva (2004). Combining Pattern Classifiers: Methods and Algorithms. John Wiley and Sons, Inc..

[41] Aha, David W., Dennis Kibler, and Marc K. Albert. "Instance-based learning algorithms." Machine learning 6.1 (1991): 37-66.

[42] J. C. Platt, Advances in Kernel Methods: Support Vector Machines, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, Dec. 1998. Fast training of support vector machines using sequential minimal optimization.

[43] Shevade, Shirish K., et al. "Improvements to the SMO algorithm for SVM regression." IEEE transactions on neural networks 11.5 (2000): 1188-1193.

[44] Keerthi, S. Sathiya, et al. "Improvements to Platt's SMO algorithm for SVM classifier design." Neural computation 13.3 (2001): 637-649.

[45] Trevor Hastie, Robert Tibshirani: Classification by Pairwise Coupling. In: Advances in Neural Information Processing Systems, 1998.

[46] Hussain, Muhammad, et al. "A comparison of SVM kernel functions for breast cancer detection." 2011 Eighth International Conference Computer Graphics, Imaging and Visualization. IEEE, 2011.

[47] J. Davis and M. Goadrich, \The Relationship Between Precision-Recall and ROC Curves," Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, 2006.

[48] Galdi, Paola, and Roberto Tagliaferri. "Data mining: accuracy and error measures for classification and prediction." Encyclopedia of Bioinformatics and Computational Biology (2018): 431-436.

[49] Powers, David Martin. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." (2011).