



## A NOVEL FRAMEWORK FOR CARDIAC ARRHYTHMIA CLASSIFICATION BASED ON SUBSPACE PROJECTION AND DECISION TREE

Semih ERGIN<sup>1</sup>, Selcan KAPLAN BERKAYA<sup>2,\*</sup>, Alper Kursat UYSAL<sup>2</sup>, Efnan SORA GUNAL<sup>3</sup>, Serkan GUNAL<sup>2</sup>, Mehmet Bilginer GULMEZOGLU<sup>1</sup>

<sup>1</sup> Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskisehir, Turkiye

<sup>2</sup> Department of Computer Engineering, Faculty of Engineering, Eskisehir Technical University, Eskisehir, Turkiye.

<sup>3</sup> Department of Computer Engineering, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Eskisehir, Turkiye

### ABSTRACT

Cardiac arrhythmia refers to abnormal activity of the heart. Correct classification of cardiac arrhythmia is, therefore, crucial for the appropriate treatment of heart diseases. In this paper, a novel approach is proposed for cardiac arrhythmia classification. Initially, the feature vectors extracted from raw electrocardiogram (ECG) signals are projected into a particular subspace obtained via the Common Vector Approach, which is an effective subspace method. The projected vectors are then fed into two distinct decision-tree-based classifiers—namely, C4.5 and random forest. The results obtained from the proposed approach are compared with those obtained with the original feature vectors using the same classifiers. For this purpose, the well-known MIT-BIH arrhythmia database was utilized. Six different sets of features based on QRS, time-domain, wavelet transform, and power spectral density are derived from ECG signals in this database. The feature sets are then used in the classification of five main beat types including non-ectopic, ventricular ectopic, supraventricular ectopic, fusion, and unknown. The experimental results reveal that the recognition performances achieved by most of the projected features are explicitly higher than those obtained with the original ones. In addition, the classification accuracy of the proposed approach climbs to 100% for the test set.

**Keywords:** Cardiac arrhythmia classification; Electrocardiography; Feature selection; Subspace projection; Decision tree

## 1. INTRODUCTION

Arrhythmia is a common name of abnormal cardiac electrical activity observation [1]. Cardiac arrhythmias frequently occur in individuals with cardiovascular diseases [2]. Many types of arrhythmias are life-threatening abnormalities that may cause cardiac arrest or sudden death. Consequently, early diagnosis of arrhythmia may extend the life and enhance the quality of life of the patient with appropriate treatment [3]. ECG is the most effective tool to diagnose arrhythmia types [2]. The role of feature extraction is vital for ECG signal analysis. In the literature, numerous works utilize various feature extraction strategies for the classification of ECG signals to detect not only arrhythmias but also other heart abnormalities. In a study, extracted RR interval series and morphology descriptors calculated from wavelet transform (WT) are used [4]. Morphological WT and time interval features [5], local binary patterns (LBP) features, and several amplitude values [6], wavelet packet entropy (WPE) features [7], convolutional neural network (CNN) features [8-10] are also studied. Also, R-peaks of ECG signals are utilized as the features [11]. Multiscale wavelet features, along with timing information, are preferred [12]. Morphology and heartbeat interval features were extracted in [13]. Time-domain (TD) features (principal components of QRS morphology, RR interval and QRS interval) are classified into four types of ECG beats—normal, ventricular ectopic, supraventricular ectopic, and fusion—in [14]. Discrete wavelet transform (DWT) [15, 16], fuzzy wavelet packet [17], tunable Q-wavelet [18] and power spectral density (PSD) [19, 20] analysis are

\*Corresponding Author: [selcankaplan@eskisehir.edu.tr](mailto:selcankaplan@eskisehir.edu.tr)

Received: 07.10.2019 Published:15.06.2020

other examples of the feature extraction approaches for ECG classification. The local fractal dimension of neighboring sample points of ECG signal segments is used as the feature in [21].

Dimension reduction also plays an essential role in ECG classification because it is useful for many different pattern recognition problems. Subspace-based feature projection methods are widely applied to reduce the dimension in most studies [1, 22-24]. In these approaches, the original feature space is projected into a new lower-dimensional subspace. An effective arrhythmia classification scheme using a dimension reduction method that combines principal component analysis (PCA) with linear discriminant analysis (LDA) is presented [23]. The experimental results show that the integration of a Probabilistic Neural Network (PNN) classifier with the proposed dimension reduction method attains satisfactory classification performance. In a previous study, three dimensionality reduction algorithms—PCA, LDA, and independent component analysis (ICA)—are individually applied to DWT sub-bands for ECG beat classification [1]. It is reported that the combination of ICA and PNN performs better than the combinations of the other dimension reduction methods and classifiers. In another study, similar methodologies are applied for the automated diagnosis of coronary artery disease [22]. Dimension reduction methods such as PCA, LDA, and ICA are applied to the set of DWT coefficients extracted from particular subbands. According to the experimental results, the combination of ICA and a Gaussian Mixture Model (GMM) classifier achieves respectable performance. In a more recent study, a new method for nonlinear feature extraction of ECG signals by combining PCA and kernel independent component analysis (KICA) is proposed [24]. This method first uses PCA to reduce the dimensions of the ECG signals and then employs KICA to calculate the feature space for extracting the nonlinear features. It is stated that the method yields satisfactory classification results with a support vector machine (SVM) classifier.

Inspired by all the approaches mentioned above, in this paper, a common vector approach (CVA)-based framework [25] is proposed for cardiac arrhythmia classification. In this framework, CVA is used to reduce the feature dimension and obtain a new subspace providing better discrimination among classes. The projected features are then fed into two distinct decision-tree-based classifiers [26]—namely, C4.5 and random forest. Although CVA has been previously used in combination with GMM for text-independent speaker recognition [27], the combination of CVA with random forest and C4.5 classifiers is newly proposed. Thus, such a combination of cardiac arrhythmia classification would be a novel approach. Besides, the impacts of different features, which have been widely preferred in the abovementioned studies, on the performance of arrhythmia classification are comparatively examined in this study. The utilized features can be briefly grouped as those based on QRS-complex, TD, WT, and PSD. While evaluating the effectiveness of the features, they are considered not only individually but also in combination. Throughout the paper, the utilized features are called CVA-projected and original, depending on whether a dimension reduction technique is applied or not, respectively. In the experiments, the well-known MIT-BIH arrhythmia database [28] was used to classify five main beat types including non-ectopic, ventricular ectopic, supraventricular ectopic, fusion, and unknown. The classification performance obtained by most of the CVA-projected features is explicitly higher than those obtained by the original ones. Also, if original features are all used together rather than individually, better performance is attained for both classifiers. However, if TD-, WT- and PSD-based features are combined, the highest success ratio is attained on CVA-projected data with the C4.5 classifier. However, when all feature sets are joined in a unique feature vector, the highest accuracy is achieved on CVA-projected data with a random forest classifier.

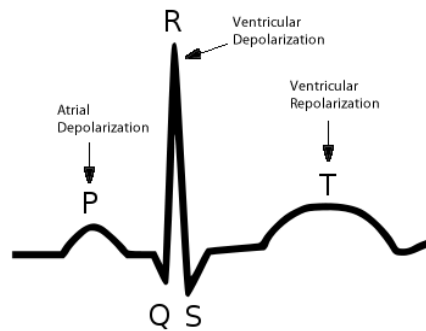
In the following sections, the utilized features are presented, the proposed method is introduced, the experimental work is explained, a discussion is given on the validity of the proposed method, and the conclusions are provided.

## 2. FEATURES

Detailed information about QRS-, TD-, WT- and PSD-based features [20, 29] is given in the following subsections. The well-known QRS detection algorithm [30] is employed to compute QRS features. Once an R peak is detected, a 200-sample-long windowing process is carried out, such that each window contains 99 samples preceding the R peak, the R peak itself, and 100 samples succeeding the R peak. TD, WT, and PSD features are then calculated over those windows.

### 2.1. QRS Features

An ECG signal typically includes five deflections—P, Q, R, S, and T waves. Because Q, R, and S waves specify a particular event, they are commonly considered together as QRS. The Q wave is the downward deflection immediately after the P wave, whereas the R wave is an upward deflection following the Q wave. Whereas the S wave is the downward deflection following the R wave, the T wave follows the S wave. The structure of the QRS complex for a sample ECG signal is illustrated in Figure 1.



**Figure 1.** A sample ECG signal

The QRS-based features used for this work are PR interval, R wave duration, P+ amplitude, QRS p-p amplitude, R wave amplitude, ST amplitude, T+ amplitude, QRS wave area, and ST slope. The efficiency of these features is proved in several studies [29, 31, 32]. The PR interval corresponds to the duration from the beginning of the P wave to the QRS complex. This specific duration includes the signals between the onset of atrial depolarization and the onset of ventricular depolarization [31]. The R wave duration indicates the time interval between the beginning and end of an R wave [32]. The P+ amplitude is the difference between the P point and the other subsequent point, where the signal rises again. The QRS p-p amplitude is computed by the difference between R and Q points in the QRS complex in terms of amplitude. The R wave amplitude specifies the height of the R wave from the baseline. The ST amplitude is the difference between the S and T points over the amplitude axis. The T+ amplitude indicates the difference between the T point and the subsequent point, where the signal rises again [32]. The QRS wave area is the area of the rectangular region over the QRS complex that is matched by using Q, R, and S points [33]. Finally, the ST slope is obtained by computing the angle of the line from the S point to the T point of the QRS complex. In this way, 9 distinct QRS features are obtained.

### 2.2. TD Features

TD features are obtained using TD values over windowed ECG signals. Specifically, eight different statistical attributes listed as energy, mean, standard deviation, maximum, minimum, kurtosis, skewness, and the difference between the maximum and minimum values of ECG signals over time frames are computed. Consequently, an 8-dimensional TD feature vector is extracted from each frame of ECG signals [20, 29].

### 2.3. WT Features

The features based on frequency information are quite successful in representing ECG signals appropriately [20, 29]. WT is an effective tool for extracting with its powerful time-frequency localization property. WT employs wavelets (scaled and shifted versions of a mother wavelet) to decompose signals into more straightforward elements.

At the start of a WT, the signal is decomposed into low- and high-frequency (approximation and detail, respectively) components. Further levels of decomposition are accomplished over the last approximation component [34]. If wavelet packet analysis is used as an alternative to WT, both approximations and details are decomposed at all levels to achieve full sub-band decomposition [35]. Because an ECG signal may carry relevant information within both its low- and high-frequency components, wavelet packet analysis rather than WT is employed to extract WT features for this work. In particular, 32 sub-bands are obtained by decomposing a 5-level wavelet packet in each frame. The energy values of each sub-band are then calculated to obtain the 32-dimensional WT feature vector per frame.

### 2.4. PSD Features

Welch's method, as mentioned in [36], is used for estimation of the PSD of an ECG signal. Following the estimation of PSD, other features can be derived easily using PSD knowledge. In this work, 11 different PSD-based features are computed. The first seven features are acquired by computing the energy, standard deviation, mean, minimum, maximum, kurtosis, and skewness values of PSD data [20, 29]. The other four features are defined as the local fractal dimension (FD) values of PSD data. The fractal, introduced by Mandelbrot [37], is a mathematical method to study self-similar structures for a long period of time. The PSD-based fractal estimator is one of the popular approaches for estimating the FD of physiological signals [21, 38]. At this step, four different FD values are computed from the equally divided parts of the PSD of the signal. It is empirically verified that dividing the PSD into more than four parts does not significantly increase the classification performance, but the processing time increases.

### 2.5. ALL Features

9 QRS, 8 TD, 32 WT, and 11 PSD features are combined to obtain 60-dimensional ALL features.

## 3. PROPOSED FRAMEWORK BASED ON SUBSPACE PROJECTION AND DECISION TREE CLASSIFICATION

In this work, a hybrid algorithm is proposed for the classification of ECG signals. The algorithm consists of a projection to a particular subspace and classification at the projected subspace. Specifically, the original feature space is first projected to a particular subspace via the CVA, which is a successful subspace method that has been previously proposed [25, 39]. The features at the projected lower-dimensional subspace are then classified using decision-tree-based classifiers—namely, C4.5 and random forest.

As the first step of CVA, eigenvalue-eigenvector pairs  $(\lambda_i, \mathbf{u}_i \mid i = 1, 2, \dots, n)$  belonging to the within-class covariance matrix  $(\Phi_j)$  of the training data are obtained, where  $j$  and  $n$  represent the class index and the dimension of the feature vector, respectively [40]. The eigenvalues are then sorted in descending order. In the next step, an  $n$ -dimensional entire feature space spanned by all eigenvectors is divided into a  $z$ -dimensional difference subspace  $(\mathbf{B}_j)$  and  $(n-z)$ -dimensional orthogonal indifference subspace  $(\mathbf{B}_j^+)$ . Consequently, the difference subspace is spanned by the eigenvectors corresponding to

the largest eigenvalues, whereas the indifference subspace is spanned by the eigenvectors corresponding to the smallest eigenvalues. The direct sum of the difference and indifference subspaces would cover the entire feature space. The motivation behind the decomposition of the entire feature space into two subspaces is to eliminate the part that has large variations from the mean [41]. In other words, the “common” information specific to the regarding class can be preserved while the variations are eliminated. Hence, class-specific information is acquired when the original space is projected onto the indifference subspace. For this computation, one may choose  $z$  such that the sum of the smallest eigenvalues is less than a particular percentage ( $L$ ) of the sum of the entire set [42] such that

$$\frac{\sum_{i=z+1}^n \lambda_i}{\sum_{i=1}^n \lambda_i} < L \quad (1)$$

If  $L$  is 0.04, good performance is usually obtained while retaining a small proportion of the variance present in the original space [43]. The value of  $z$  can also be determined using the point where the eigenvalues of the training data start varying slowly upon the plot of the eigenvalues in descending order. CVA has been previously used for both pattern classification [40, 44, 45] and feature selection [46, 47] purposes.

Decision trees can be described as multistage decision systems in which classes are consecutively rejected until an accepted class is found [26, 48]. In these systems, the feature space is split into unique sections corresponding to the classes. One of the well-known decision trees is the binary classification tree, which splits the feature space sequentially into two sections by comparing feature values with a particular threshold. In this way, an unknown feature vector is assigned to a specific class through a sequence of yes/no decisions through a path of the nodes of the tree. When designing a classification tree, the splitting criterion, stop-splitting rule, and class assignment rule should be carefully considered. The main purpose of splitting the feature space is to obtain subsets, which are more class homogeneous than the former subsets. Hence, the splitting criterion at any node is to find the split that provides the highest possible reduction in node impurity. If the highest reduction in node impurity is smaller than a particular threshold or a single class is acquired following a split, then the splitting process is ended. If a node is appointed to be a leaf or terminal, then a class assignment is made. A common assignment method is to use majority rule that assigns a leaf to the class to which the majority of the feature vectors in the corresponding subset belong. C4.5 [49] is one of the widely used decision-tree-based classification algorithms. Similarly, the random forest classifier is another popular approach that operates by constructing an ensemble of multiple decision trees [50].

The motivation behind the proposed method can be briefly described as follows: As previously mentioned, the original space is projected onto the indifference subspace via CVA. Because the indifference subspace is attained by the eigenvectors corresponding to the smallest eigenvalues, the variations among class samples are minimized at this subspace. As a consequence, class homogeneity increases substantially with respect to the original space. Because the decision-tree-based methods classify the homogeneously distributed data much better, the classification with either C4.5 or random forest classifier at the indifference subspace would provide higher performance than that of the classification in the original space.

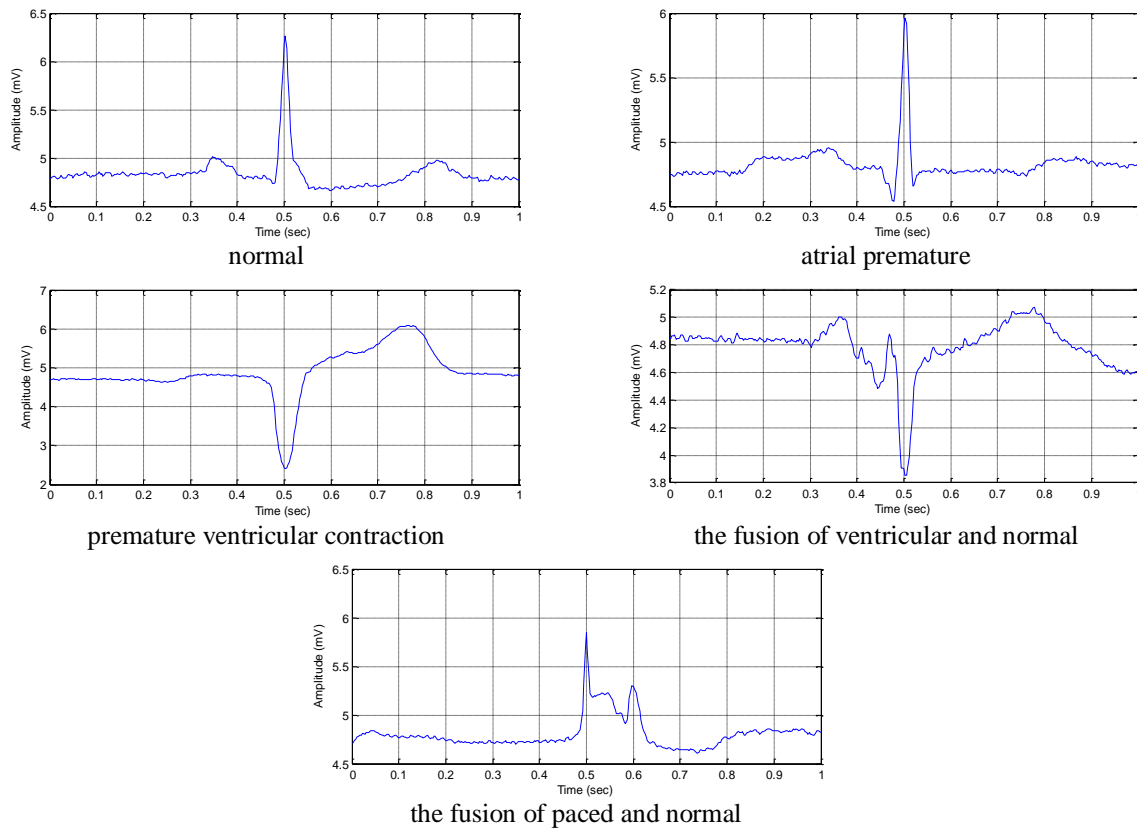
#### 4. EXPERIMENTS

In this work, six sets of features were classified with two distinct decision-tree-based classifiers—namely, C4.5 and random forest. The feature sets mentioned previously were employed both individually and in combination within the classification framework. In particular, QRS features; TD features; WT features; PSD features; the combination of TD, WT, and PSD features (TD+WT+PSD); and finally, the combination of all four feature sets (ALL features) are evaluated. This evaluation was realized on these six types of features and their projected versions obtained using CVA. The success

measures of the classification performance were chosen as accuracy (Acc), positive predictivity (P+), and sensitivity (Sen) values. Throughout the experiments, the 10-fold cross-validation was used for a fair assessment. Details of the database and the results of the experimental analysis are presented in the following subsections.

#### 4.1. ECG Database

As mentioned previously, the entire MIT-BIH arrhythmia database [28] was used to conduct the experiments. The database includes ECG recordings obtained from 47 subjects studied by Boston's Beth Israel Hospital Arrhythmia Laboratory. The categorization of ECG signals within the database was carried out considering the ANSI/AAMI EC57:1998 standard [51, 52]. According to this standard, there are 15 distinct beat types grouped under five main classes: non-ectopic (N), ventricular ectopic (V), supraventricular ectopic (S), fusion (F), and unknown (Q). The distributions of those classes and the numbers of beats in each class are given in Table 1. Sample ECG signals from each of these five classes are also provided in Figure 2. In the literature, numerous works follow the AAMI recommendation on the categorization of the beats described above; [1, 53-55] are just some recent examples of those works.



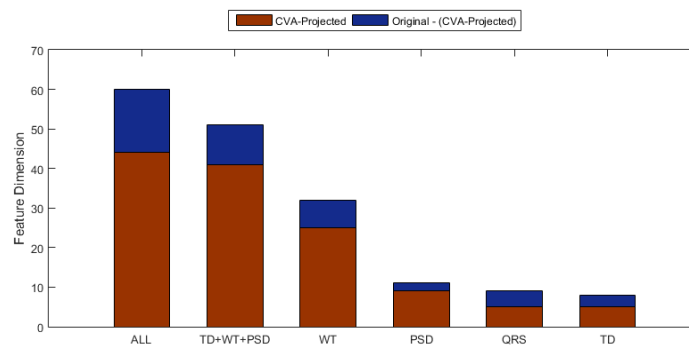
**Figure 2.** Sample ECG signals from the MIT-BIH arrhythmia database

**Table 1.** Contents of the ECG database

Symbol	Beat Class	Subcategory	# of beats
N	Non-Ectopic	Normal	71.349
		Left Bundle Branch Block	
		Right Bundle Branch Block	
		Atrial Escape	
		Nodal Escape	
S	Supraventricular Ectopic	Atrial Premature	1.862
		Aberrated Atrial Premature	
		Nodal Premature	
V	Ventricular Ectopic	Supraventricular Premature	3.008
		Premature Ventricular Contraction	
F	Fusion	Ventricular Escape	559
Q	Unknown	Fusion of Ventricular and Normal	
		Paced	
		Fusion of Paced and Normal	
		Unclassifiable	

#### 4.2. Dimension Reduction Analysis

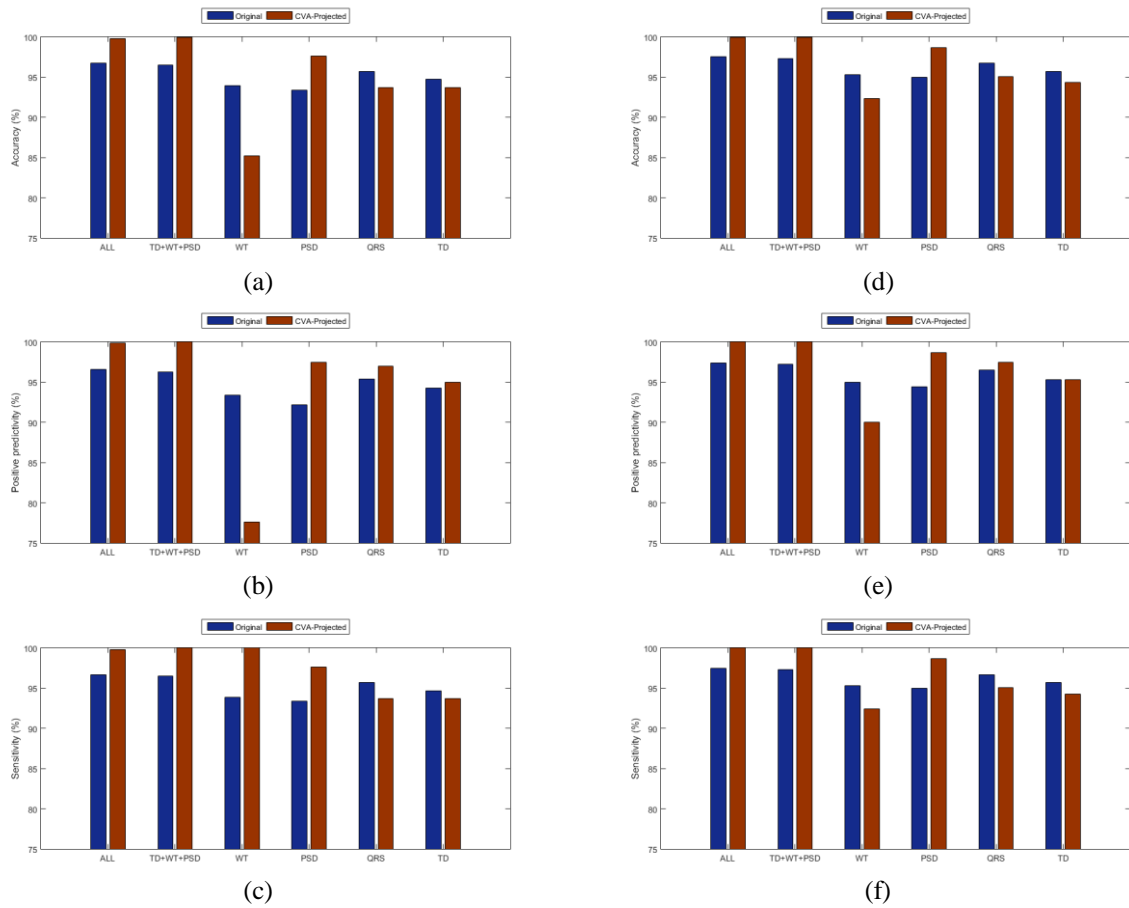
In this study, CVA is used to reduce the dimension and increase the performance of classification. The dimension reduction rate is shown in Figure 3. In this figure, the blue bars represent the differences in dimensions between the original and CVA-projected feature sets. The dimensions of the original feature set were 60, 51, 32, 11, 9, and 8 for ALL, TD+WT+PSD, WT, PSD, QRS, and TD, respectively. According to Figure 3, the dimension reduction ratio varies from 0.22 to 0.44 for different feature sets.

**Figure 3.** Comparison of original and CVA-projected feature dimensions

#### 4.3. Remarks on Recognition Results

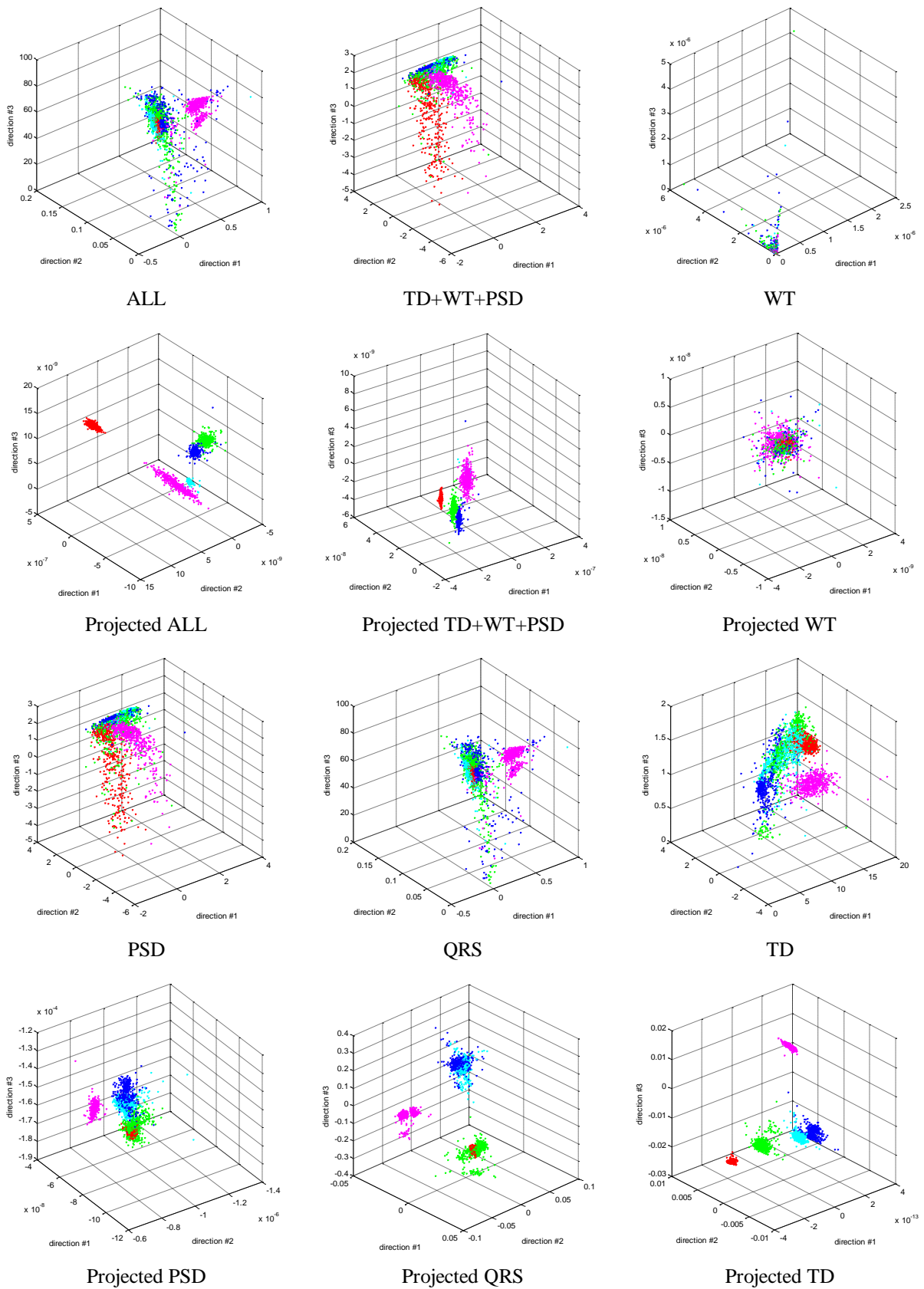
The performance of arrhythmia classification schemes utilizing the feature sets is summarized in Figure 4. According to this figure, the classification performance of CVA-projected data is higher than the performance of the original feature sets in all cases except for WT, QRS, and TD features. Also, the random forest classifier is more successful than the C4.5 classifier for both original and CVA-projected feature sets. One can note from this figure that the highest accuracy score was found to be nearly 100% when ALL or (TD+WT+PSD) feature sets projected by CVA are used with either the C4.5 or random forest classifier. Conversely, the worst results are achieved mostly with WT features. This situation can also be visually observed in Figure 5.

Considering those outcomes, because the best results are achieved without QRS based features, one may also state that the detection of QRS, which is a complicated and time-consuming process, is not mandatory at all to be able to achieve the best results.



**Figure 4.** Classification results in terms of Acc, P+, Sen, (a-c) C4.5, (d-f) Random forest





**Figure 5.** Plots of 3-D feature vectors for the original and projected feature sets (each class is plotted using a different color)

#### 4. DISCUSSIONS

In this paper, a novel classification scheme is proposed, and its success on a famous MIT-BIH arrhythmia database is elaborately investigated. The proposed scheme obviously increases the recognition performance for different sets of features. A thorough explanation of this performance improvement was attempted using two different approaches.

The first approach is based on the values that indicate the separability of the classes. These values are obtained from the following inequality:

$$\|m^i - m^j\| > \max \{\|x_k^i - m^i\|\} + \max \{\|x_k^j - m^j\|\} \quad (2)$$

In this inequality,  $m^i$  and  $m^j$  are the average feature vectors of the  $i$ th and  $j$ th classes, respectively, whereas  $x_k^i$  and  $x_k^j$  stand for the  $k$ th feature vectors of the  $i$ th and  $j$ th classes, respectively. The distance between the means of two classes is compared with the summation of the maximum distance between the farthest feature vector in the  $i$ th class and the average of the  $i$ th class and the maximum distance between the farthest feature vector in the  $j$ th class and the average of the  $j$ th class. Two different distance values are computed using this equation, and they are listed in Table 2 for all feature types. In this table, the first row of each feature set indicates the distances between the corresponding class averages, whereas the second row shows the summation values of the maximum distances. The values in the first rows are computed using

$$\|m^i - m^j\| \quad (3)$$

whereas the values in the second rows are calculated by

$$\max \{\|x_k^i - m^i\|\} + \max \{\|x_k^j - m^j\|\} \text{ for } k = 1, 2, \dots, N \quad (4)$$

where  $N$  indicates the total number of feature vectors in each class. For better classification, the distance value computed by Eq. (3) should be as large as possible with respect to the one computed by Eq. (4). As shown in Table 2, this condition is true (indicated in bold) for the feature sets (ALL) and (TD+WT+PSD). This outcome implies that the CVA-based projection not only reduces the dimension of the feature vectors but also provides better class discrimination for appropriate feature sets.

The second approach provides visual proof of the effectiveness of the proposed method. The three-dimensional feature vectors calculated by the projection of feature vectors onto the eigenvectors corresponding to the top three eigenvalues are plotted for each of twelve feature vector conditions. All of these cases are illustrated in Figure 5. The usage of the eigenvectors corresponding to the top three eigenvalues for the projection offers the most difficult circumstances to visually separate the classes because the eigenvectors corresponding to the highest eigenvalues determine the directions where data distributions have the largest possible variances—that is, as much variability in the data as possible. Although this hardest situation is experienced, the classes can be visually distinguished except for WT features. Therefore, one can easily infer from those figures that the CVA-based dimension reduction process selects the features reserving the most discriminative information. This situation is consistent with the values given in Figure 4 such that the projected versions of ALL, TD+WT+PSD, and PSD feature sets increase the classification performance.

In addition to these two different approaches, the results attained in this work are compared with those given in various published works in the literature. The comparison is carried out in terms of database size, the number of classes, types of features, utilized classifiers, and the three different success metrics of Acc, P+, and Sen. The outcome of the comparison is provided in Table 3. In this table, the

best results of the proposed work are separately listed for C4.5 and random forest classifiers. One can easily see that the proposed work surpasses the other references in terms of Acc, P+, and Sen values.

**Table 2.** Distance values for the original and projected feature sets

Feature Set	Equation	[C1,C2]	[C1,C3]	[C1,C4]	[C1,C5]	[C2,C3]	[C2,C4]	[C2,C5]	[C3,C4]	[C3,C5]	[C4,C5]
ALL	Eq.(3)	0.05	0.38	0.38	1.58	0.38	0.38	1.60	0.02	1.29	1.28
	Eq.(4)	22.32	22.90	22.57	35.25	31.04	30.71	43.39	31.29	43.97	43.64
Projected ALL	Eq.(3)	2.44	1.86	1.83	2.58	1.53	2.34	1.20	2.59	1.83	2.23
	Eq.(4)	<b>1.52</b>	2.39	<b>1.41</b>	<b>1.51</b>	<b>1.20</b>	<b>0.22</b>	<b>0.32</b>	<b>1.10</b>	<b>1.19</b>	<b>0.21</b>
TD+WT+PSD	Eq.(3)	1.82	2.49	1.15	3.54	4.01	2.81	4.07	1.54	4.85	4.16
	Eq.(4)	16.41	14.61	15.23	18.20	11.24	11.86	14.83	10.06	13.03	13.65
Projected TD+WT+PSD	Eq.(3)	1.53	1.73	1.81	1.11	1.29	1.16	0.69	0.98	1.60	1.41
	Eq.(4)	<b>1.52</b>	1.91	<b>1.35</b>	1.73	<b>0.78</b>	<b>0.22</b>	<b>0.60</b>	<b>0.62</b>	<b>0.99</b>	<b>0.43</b>
WT	Eq.(3)	1.75	2.65	1.11	2.44	4.11	2.76	3.73	1.60	3.70	2.59
	Eq.(4)	7.05	7.87	8.07	13.71	3.26	3.46	9.10	4.29	9.93	10.13
Projected WT	Eq.(3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Eq.(4)	0.04	0.04	0.03	0.08	0.01	0.01	0.05	0.01	0.05	0.05
PSD	Eq.(3)	0.90	0.56	0.57	2.92	1.13	1.37	2.11	0.38	3.06	3.37
	Eq.(4)	7.42	8.33	4.71	3.81	1.31	9.50	8.60	10.41	9.51	5.88
Projected PSD	Eq.(3)	4.98	7.09	6.12	5.28	6.92	4.72	4.26	4.38	4.62	1.02
	Eq.(4)	6.38	11.31	7.09	10.14	7.62	<b>3.40</b>	6.45	8.33	11.38	7.16
QRS	Eq.(3)	3.59	1.38	3.23	1.11	4.88	0.37	2.70	4.53	2.26	2.37
	Eq.(4)	18.32	24.15	18.99	34.12	28.55	23.38	38.51	29.21	44.35	39.18
Projected QRS	Eq.(3)	0.25	0.30	0.17	0.15	0.24	0.25	0.29	0.31	0.21	0.18
	Eq.(4)	1.68	1.84	2.11	2.00	0.75	1.03	0.91	1.18	1.07	1.34
TD	Eq.(3)	1.90	2.22	0.44	2.22	3.82	2.16	3.41	1.82	3.57	2.50
	Eq.(4)	13.86	16.31	14.20	17.02	12.12	10.01	12.82	12.46	15.27	13.16
Projected TD	Eq.(3)	0.04	0.10	0.04	0.10	0.10	0.04	0.10	0.11	0.06	0.13
	Eq.(4)	0.42	0.35	0.42	0.52	0.34	0.41	0.51	0.34	0.44	0.51

**Table 3.** Comparison of the proposed work to the literature.

Reference	Classes	Features	Classifier	Acc	P+	Sen
[1]	5	WT	SVM, NN and PNN	99.28	99.21	97.97
[5]	5	TD, WT	ANN	95.58	61.48	62.74
[6]	4	TD, WT, LBP	SVM	94.50	66.40	70.30
[7]	5	TD, WPE	KNN, DT, PNN, SVM, RF	94.61	-	-
[8]	5	CNN	CNN	89.07	63.14	95.90
[9]	5	CNN	CNN	96.60	71.67	67.37
[13]	5	QRS	LDC	99.20	98.40	94.40
[15]	5	TD, WT	SVM-RBF, NN	98.91	-	98.91
[53]	5	TD	Decision tree	96.18	96.15	96.16
[54]	5	TD	BBNN-PSO	97.00	-	-
[56]	3	TD, WT, QRS	LDC, QDC	93.00	74.67	84.33
[57]	5	QRS	Dynamic clustering	98.84	91.67	95.00
[58]	5	Raw wave	Reservoir Computing	98.43	88.75	84.83
[59]	5	PCANet	SVM, KNN, BPNN, RF	97.77	95.32	86.35
Proposed work	5	TD, WT, PSD	C4.5	99.98	100.00	100.00
Proposed work	5	TD, WT, PSD	Random forest	99.99	100.00	100.00

## **5. CONCLUSIONS**

In this paper, the well-known MIT-BIH arrhythmia database was used to classify five main beat types: non-ectopic beats, supraventricular ectopic beats, ventricular ectopic beats, fusion beats, and unknown beats. For this purpose, six different sets of features (QRS, TD, WT, PSD, TD+WT+PSD, and ALL) were derived from ECG signals in this database. In the proposed approach, these original features are first projected onto the indifference subspace formed by CVA so that a dimension reduction process is applied to the original features. Second, the projected feature sets are treated with two distinct decision-tree-based classifiers—namely, C4.5 and random forest. This combination of CVA with the random forest and C4.5 classifiers is newly applied to arrhythmia beat classification. The experimental results clearly show that the performance attained with most of the CVA-projected features is higher than the performance of the original features. Because the indifference subspace is attained by the eigenvectors corresponding to the smallest eigenvalues, the variations among class samples are minimized in this subspace. As a consequence, class homogeneity increases substantially with respect to the original space. This situation can be observed in Figure 5 and Table 2, as well. Because the decision-tree-based methods classify the homogeneously distributed data much better, the classification with either the C4.5 or random forest classifier in the indifference subspace provides higher performance than that of the classification in the original space. The results of the experimental work summarized in Figure 4 also verify this statement.

Comparatively considering the two classifiers, the random forest classifier is more successful than the C4.5 classifier for both original and CVA-projected feature sets. It has been experimentally verified that better performance is achieved for both classifiers if all original features are used not individually but together instead. However, the original feature sets TD+WT+PSD and ALL have similar performance; that is, QRS features do not affect the success of the classification noticeably if the remaining features are used.

CVA-projected feature sets offer classification accuracies higher than 85–90% in general. These values may increase to 100% in some cases. One can note in Figure 4d that the highest classification accuracy was found to be nearly 100% when the projected ALL or TD+WT+PSD features are used with either the C4.5 or random forest classifier.

Finally, the proposed approach considerably increases the performance obtained from the original feature sets because the dimension reduction provided by CVA projection exposes the most discriminative features.

As future work, subspace projection methods other than CVA (e.g., PCA and Fisher's LDA) can be also employed to investigate the efficiency of new features. The contribution of feature selection methods can be analyzed rather than projecting features into particular subspaces. Additional classification algorithms can be tested. The proposed approach can be applied to other ECG-related studies such as automatic screening of obstructive sleep apnea cases as well.

## **ACKNOWLEDGEMENTS**

This work was supported by Eskişehir Osmangazi University, Fund of Scientific Research Projects under grant number 201215037.

## **REFERENCES**

- [1] Martis RJ, Acharya UR, Min LC. ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform. *Biomed Signal Proces* 2013; 8: 437-448. doi:10.1016/j.bspc.2013.01.005.

- [2] Kaplan Berkaya S, Uysal AK, Sora Gunal E, et al. A survey on ECG analysis. *Biomed Signal Proces* 2018; 43: 216-35. doi:10.1016/j.bspc.2018.03.003.
- [3] Haseena HH, Mathew AT, Paul JK. Fuzzy clustered probabilistic and multi layered feed forward neural networks for electrocardiogram arrhythmia classification. *J Med Syst* 2011; 35: 179-188. doi:10.1007/s10916-009-9355-9.
- [4] Llamedo M, Martinez JP. An automatic patient-adapted ECG heartbeat classifier allowing expert assistance. *IEEE T Bio-Med Eng* 2012; 59: 2312-2320. doi:10.1109/Tbme.2012.2202662.
- [5] Ince T, Kiranyaz S, Gabbouj M. A generic and robust system for automated patient-specific classification of ECG signals. *IEEE T Bio-Med Eng* 2009; 56: 1415-1426. doi:10.1109/Tbme.2009.2013934.
- [6] Mondéjar-Guerra V, Novo J, Rouco J, et al. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomed Signal Proces* 2019; 47: 41-48. doi: 10.1016/j.bspc.2018.08.007.
- [7] Li T, Zhou M. ECG classification using wavelet packet entropy and random forests. *Entropy* 2016; 18: 285. doi: 10.3390/e18080285.
- [8] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, San Tan, R. A deep convolutional neural network model to classify heartbeats. *Comput Biol Med* 2017; 89: 389-396. doi: 10.1016/j.compbimed.2017.08.022.
- [9] Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE T Bio-Med Eng* 2016; 63: 664-675. doi: 10.1109/TBME.2015.2468589.
- [10] Chen C, Hua Z, Zhang R, et al. Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomed Signal Proces* 2020; 57: 101819. doi:10.1016/j.bspc.2019.101819.
- [11] Jiang W, Kong SG. Block-based neural networks for personalized ECG signal classification. *IEEE T Neural Networ* 2007; 18: 1750-1761. doi:10.1109/Tnn.2007.900239.
- [12] Inan OT, Giovangrandi L, Kovacs GTA. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE T Bio-Med Eng* 2006; 53: 2507-2515. doi:10.1109/TBME.2006.880879.
- [13] de Chazal P, Reilly RB. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. *IEEE T Bio-Med Eng* 2006; 53: 2535-2543. doi:10.1109/TBME.2006.883802.
- [14] Hu YH, Palreddy S, Tompkins WJ. A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE T Bio-Med Eng* 1997; 44: 891-900. doi:10.1109/10.623058.
- [15] Elhaj FA, Salim N, Harris AR, et al. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput Meth Prog Bio* 2016; 127: 52-63. doi:10.1016/j.cmpb.2015.12.024.

- [16] Sangaiah AK, Arumugam M, Bian GB. An intelligent learning approach for improving ECG signal classification and arrhythmia analysis. *Artif Intell Med* 2020; 103: 101788. doi:10.1016/j.artmed.2019.101788.
- [17] Khushaba RN, Kodagoda S, Lal S, et al. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE T Bio-Med Eng* 2011; 58: 121-131. doi:10.1109/Tbme.2010.2077291.
- [18] Jha CK, Kolekar MH. Cardiac arrhythmia classification using tunable Q-wavelet transform based features and support vector machine classifier. *Biomed Signal Proces* 2020; 59: 101875. doi:10.1016/j.bspc.2020.101875.
- [19] Ubeyli ED. Statistics over features of ECG signals. *Expert Syst Appl* 2009; 36: 8758-8767. doi:10.1016/j.eswa.2008.11.015.
- [20] Gunal S, Ergin S, Gunal ES, et al. ECG classification using ensemble of features. In: *IEEE 47th Annual Conference on Information Sciences and Systems*; 20-22 March 2013; Baltimore, MD, USA: IEEE, 1-5. doi:10.1109/CISS.2013.6624256.
- [21] Mishra AK, Raghav S. Local fractal dimension based ECG arrhythmia classification. *Biomed Signal Proces* 2010; 5: 114-123. doi:10.1016/j.bspc.2010.01.002.
- [22] Giri D, Acharya UR, Martis RJ, et al. Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform. *Knowl-Based Syst* 2013; 37: 274-282. doi:10.1016/j.knosys.2012.08.011.
- [23] Wang JS, Chiang WC, Hsu YL, Yang YTC. ECG ECG arrhythmia classification using a probabilistic neural network with a feature reduction method. *Neurocomputing* 2013; 116: 38-45. doi:10.1016/j.neucom.2011.10.045.
- [24] Li H, Liang H, Miao C, et al. Novel ECG signal classification based on KICA nonlinear feature extraction. *Circuits, Systems, and Signal Processing* 2016; 35: 1187-1197. doi:10.1007/s00034-015-0108-3.
- [25] Gulmezoglu MB, Dzhafarov V, Barkana A. The common vector approach and its relation to principal component analysis. *IEEE Transactions on Speech and Audio Processing* 2001; 9: 655-662. doi:10.1109/89.943343.
- [26] Theodoridis S. and Koutroumbas K. *Pattern Recognition*. 4th ed. Academic Press, 2009.
- [27] Sadic S, Gulmezoglu MB. Common vector approach and its combination with GMM for text-independent speaker recognition. *Expert Syst Appl* 2011; 38: 11394-11400. doi:10.1016/j.eswa.2011.03.009.
- [28] Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet - Components of a new research resource for complex physiologic signals. *Circulation* 2000; 101: E215-E220.
- [29] Ergin S, Uysal A, Gunal, ES, et al. ECG based biometric authentication using ensemble of features. In: *IEEE 9th Iberian Conference on Information Systems and Technologies*; 18-21 June 2014; Barcelona, Spain. doi:10.1109/CISTI.2014.6877089.

- [30] Pan J, Tompkins WJ. A real-time QRS detection algorithm. *IEEE T Bio-Med Eng* 1985; 32: 230-236. doi: 10.1109/Tbme.1985.325532.
- [31] Valentin JP. Reducing QT liability and proarrhythmic risk in drug discovery and development. *Brit J Pharmacol* 2010; 159: 5-11. doi:10.1111/j.1476-5381.2009.00547.x.
- [32] Biel L, Pettersson O, Philipson L, et al. ECG analysis: A new approach in human identification. *IEEE T Instrum Meas* 2001; 50: 808-812. doi: 10.1109/19.930458.
- [33] Stravroulakis P, Stamp M. *Handbook of Information and Communication Security*. Springer, 2010.
- [34] Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press, 2001.
- [35] Coifman RR, Wickerhauser MV. Entropy-based algorithms for best basis selection. *IEEE T Inform Theory* 1992; 38: 713-718. doi: 10.1109/18.119732.
- [36] Welch PD. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 1967; 15: 70-73. doi:10.1109/TAU.1967.1161901.
- [37] Mandelbrot BB. *The Fractal Geometry of Nature*. W. H. Freeman and Company, 1983.
- [38] Spasic S. Spectral and fractal analysis of biosignals and coloured noise. In: *IEEE 5th International Symposium on Intelligent Systems and Informatics*; 24-25 August 2007; Subotica, Serbia. IEEE, 147-149. doi:10.1109/SISY.2007.4342641.
- [39] Gulmezoglu MB, Keskin M, Dzhafarov V, et al. A novel approach to isolated word recognition. *IEEE Transactions on Speech and Audio Processing* 1999; 7: 620-628. doi:10.1109/89.799687.
- [40] Gulmezoglu MB, Ergin S. An approach for bearing fault detection in electrical motors. *European Transactions on Electrical Power* 2007; 17: 628-641. doi:10.1002/etep.161.
- [41] Landgrebe D. Hyperspectral image data analysis. *IEEE Signal Processing Magazine* 2002; 19: 17-28. doi:10.1109/79.974718.
- [42] Oja E. *Subspace Methods of Pattern Recognition*. John Wiley and Sons, 1983.
- [43] Swets DL, Weng J. Using discriminant eigenfeatures for image retrieval. *IEEE T Pattern Anal* 1996; 18: 831-836. doi:10.1109/34.531802.
- [44] Cevikalp H, Neamtu M, Wilkes M, et al. Discriminative common vectors for face recognition. *IEEE T Pattern Anal* 2005; 27: 4-13. doi:10.1109/TPAMI.2005.9.
- [45] Yumurtaci M, Gokmen G, Arıkan C, et al. Classification of short circuit faults in high voltage energy transmission line using energy of instantaneous active power components based common vector approach. *Turk J Electr Eng Co* 2016; 24: 1901-1915. doi:10.3906/elk-1312-131.
- [46] Gunal S, Ergin S, Gulmezoglu MB, et al. On feature extraction for spam e-mail detection. *Lecture Notes on Computer Science* 2006; 4105: 635-642. doi:10.1007/11848035\_84.

- [47] Gunal S, Edizkan R. Subspace based feature selection for pattern recognition. *Inform Sciences* 2008; 178: 3716-3726. doi:10.1016/j.ins.2008.06.001.
- [48] Gunal S. Hybrid feature selection for text classification. *Turk J Electr Eng Co* 2012; 20: 1296-1311. doi:10.3906/elk-1101-1064.
- [49] Quinlan JR. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [50] Ho TK. Random decision forests. In: *IEEE Proceedings of the Third International Conference on Document Analysis and Recognition*; 14-16 August 1995; Montreal, Quebec, Canada. IEEE, 278-282. doi:10.1109/ICDAR.1995.598994.
- [51] ANSI/AAMI EC57: Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms (AAMI Recommended Practice/American National Standard), Order Code: EC57-293, 1998.
- [52] AAMI Recommendation, Recommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms, Association for the Advancement of Medical Instrumentation, 1987.
- [53] Afkhami RG, Azarnia G, Tinati MA. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. *Pattern Recogn Lett* 2016; 70: 45-51. doi:10.1016/j.patrec.2015.11.018.
- [54] Shadmand S, Mashoufi B. A new personalized ECG signal classification algorithm using Block-based Neural Network and Particle Swarm Optimization. *Biomed Signal Proces* 2016; 25: 12-23. doi:10.1016/j.bspc.2015.10.008.
- [55] Queiroz V, Luz E, Moreira G, et al. Automatic cardiac arrhythmia detection and classification using vectorcardiograms and complex networks, In: *IEEE 37th Annual International Conference of the Engineering in Medicine and Biology Society*; 25-29 August 2015; Milan, Italy. IEEE, 5203-5206. doi:10.1109/EMBC.2015.7319564.
- [56] Llamedo M, Martinez JP. Heartbeat classification using feature selection driven by database generalization criteria. *IEEE T Bio-Med Eng* 2011; 58: 616-625. doi:10.1109/Tbme.2010.2068048.
- [57] Castro D, Félix P, Presedo J. A method for context-based adaptive QRS clustering in real time. *IEEE J Biomed Health* 2015; 19: 1660 - 1671. doi:10.1109/JBHI.2014.2361659.
- [58] Escalona-Morán MA, Soriano MC, Fischer I, et al. Electrocardiogram classification using reservoir computing with logistic regression. *IEEE J Biomed Health* 2015; 19: 892-898. doi: 10.1109/JBHI.2014.2332001.
- [59] Yang W, Si Y, Wang D, Guo B. Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. *Comput Biol Med* 2018; 101: 22-32. doi: 10.1016/j.compbimed.2018.08.003.