

RESEARCH ARTICLE

Heartbeat type classification with optimized feature vectors

Özal Yıldırım*, Ulas Baran Baloglu

Department of Computer Engineering, Munzur University, Turkey
oyildirim@munzur.edu.tr, ulasbaloglu@gmail.com

ARTICLE INFO

Article history:

Received: 16 December 2017

Accepted: 7 March 2018

Available Online: 12 April 2018

Keywords:

ECG signals

Feature optimization

Feature vectors

Classification

Machine learning

AMS Classification 2010:

68TXX, 62-07, 92-08

ABSTRACT

In this study, a feature vector optimization based method has been proposed for classification of the heartbeat types. Electrocardiogram (ECG) signals of five different heartbeat type were used for this aim. Firstly, wavelet transform (WT) method were applied on these ECG signals to generate all feature vectors. Optimizing these feature vectors is provided by performing particle swarm optimization (PSO), genetic search, best first, greedy stepwise and multi objective evolutionary algorithms on these vectors. These optimized feature vectors are later applied to the classifier inputs for performance evaluation. A comprehensive assessment was presented for the determination of optimized feature vectors for ECG signals and best-performing classifier for these optimized feature vectors was determined.



1. Introduction

Electrocardiogram (ECG) signals have a critical importance for determining abnormal heart conditions. Computer-aided analysis of these data is particularly considerable regarding the development of smart medical platforms. For this purpose, it is necessary to automatically detecting and recognizing heartbeats from ECG signals.

Heartbeat detection process from the ECG signals is not easy to realize because of the various types of noise, which exist in the ECG signals [1]. Different techniques have been comprehensively analyzed for beat detection such as digital filter usage [2, 3], the wavelet transform (WT) [4-7], neural networks (NN) [8, 9], hidden markov model [10] and particle swarm optimization (PSO) [11]. Detected heartbeat signals are segmented to be used in the classification systems.

There are two main stages in the ECG signal classification. These are feature extraction and classification stages. In the feature extraction stage, distinguishing features of ECG signals are revealed. These features are used to generate feature vectors for each signal. Feature extraction is required to remove unnecessary, noisy or corrupted inputs. This stage improves the accuracy of classifiers which are used in the consequent classification stage. In the classification stage, a suitable classifier was trained on the obtained features.

WT methods such as multiresolution and discrete types [12-16] are frequently used for feature extraction process. There are also other studies where higher order statistic [17] and mathematical morphology method [18] were employed. In the classification stage, different classifiers can be applied such as extreme learning machines [19] and support vector machines [20-22]. NN classifier also was widely used in classification studies [23-25].

In this study, feature vector optimization approach was used to classify heartbeat signals. The PSO search [26], genetic search [27], best first-greedy stepwise [28] and multi-objective evolutionary search [29] methods are applied on these feature vectors for reducing the computational complexity of the overall process. The optimized vectors were passed as inputs to the classifier algorithms. Random forests (RF) [30] and least square support vector machines (LS-SVM) [31] classifiers were selected for determining the heartbeat types. Experimental results were conducted on an ECG dataset which includes five different heartbeat type. These ECG signals are collected from the MIT-BIH arrhythmia database.

The rest of this paper is organized as follows: Section 2 briefly presents the feature optimization problem. Section 3 describes the proposed method in detail. Section 4 presents the experimental results and Section 5 concludes the paper.

*Corresponding author

2. Feature optimization

Feature optimization is one of the most significant challenges in data analysis studies due to the huge volume of data to be processed. Feature selection and optimization reduce the dimension of the data by removing unnecessary features so that these processes improve the performance of algorithms. There are variously supervised, semi-supervised and unsupervised feature optimization techniques in the literature.

The main idea of this optimization process is gathering a subset of existing features by eliminating features which are containing relatively little information. The relevance relation of a set and the target class should be defined to facilitate the feature optimization. Let X set denotes the input features and Y set denotes the relevant classes.

$$X = \{x_1, x_2, \dots, x_m\}, Y = \{y_1, y_2, \dots, y_n\} \quad (1)$$

Among $\{x, y\}$ pairs, the objective function of feature selection is finding a subset of pairs which can be defined as follows,

$$S(\bar{X}) = \max_{|Y| < m, (Y \subset X)} O(Y) \quad (2)$$

In equation (2), O function is the feature optimization function which calculates the accuracy of a feature subset. In feature optimization, the search space contains all the possible subsets of features so that feature optimization can significantly affect the performance of ECG signal classification. Consequently, feature selection and optimization problem is an NP-hard problem [32] so that meta-heuristics such as evolutionary algorithms are frequently considered in creating a solution space when there is a large number of features [33]. It is essential to determine the best techniques for specific tasks such as beat detection and recognition. In this study, we analyze ECG signal with various optimization methods and classifier algorithms.

3. The proposed method

3.1 Feature vectors

The first step in the process of determining the optimal feature vectors for ECG signals is to make feature deductions on these signals. For this purpose, 6 level Dabuchies (Db6) [34] wavelet transform method was used in the study. By using wavelet transform, coefficient matrices are obtained for the signals separated into lower frequency bands. The steps of this process are shown in detail in Figure 1.

In the wavelet transform process, input signals are passed through high-pass and low-pass filters after each conversion. These filters provide a detailed analysis of high and low-frequency components of the signal. As a result of the wavelet transform, approximation (A^n) and detail (D^n) coefficients are formed at each level of the input signal.

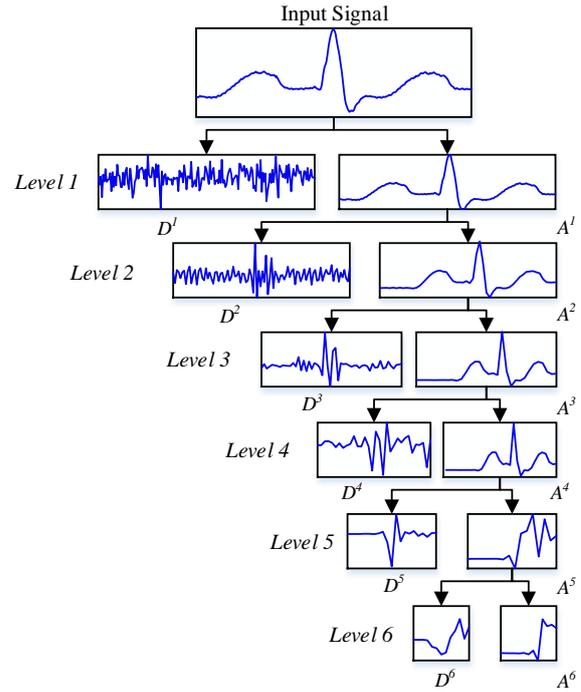


Figure 1. Wavelet transform steps for 6 levels.

Coefficient matrices do not have an appropriate use because of the large size data classifiers they contain. For this reason, data in these coefficient matrices have to be reduced to lower dimensional representing data. Statistical methods were commonly employed for this purpose. In this study, energy, mean, standard deviation and norm entropy methods were used. Energy calculation on coefficient matrices can be defined as:

$$E^i = \sum_{k=1}^N (C_k^i)^2 \quad \text{for } i=1 \dots M \quad (3)$$

where C denotes coefficient matrices, and N is the size of these matrices. Finally, M denotes the number of the sub-bands. The average of coefficient matrices is calculated as follows,

$$\mu^i = \frac{1}{N} \sum_{k=1}^N C_k^i \quad (4)$$

Thus, an average value is obtained for the coefficient matrices of each frequency sub-band. Another employed method, standard deviation calculation, is as follows,

$$\sigma^i = \sqrt{\frac{1}{N} \sum_{k=1}^N (C_k^i - \mu^i)^2} \quad (5)$$

Lastly, the norm entropy calculation for each coefficient matrix is obtained by the following equation.

$$\rho^i = \sum_{k=1}^N (C_k^i)^p \quad 1 \leq p \quad (6)$$

The coefficient matrix for each input signal consists of a total of 7 coefficient matrices as:

$$C^i = [D_1, D_2, D_3, D_4, D_5, D_6, A_6]$$

All of these calculations generate a 28-dimensional feature vector for each signal. This 28-dimensional feature vector can be defined as:

$$V^k = [E_j^k, \mu_j^k, \sigma_j^k, \rho_j^k] \text{ for } j=1 \dots 7 \text{ and } k=1 \dots N \quad (7)$$

where k is number of feature vectors, and j is number of sub-bands respectively.

3.2. ECG dataset

MIT-BIH arrhythmia [35] database, which is widely used in the literature, was selected as ECG data source in experimental studies. Within this database, there are 48 different records and beats of these records were labelled by experts. Normal, Premature Ventricular Contraction (PVC), Paced, Left Bundle Branch Block (LBBB), and Right Bundle Branch Block (RBBB) heartbeats are selected from the data. In Figure 2 the signal form of a normal heartbeat is given.

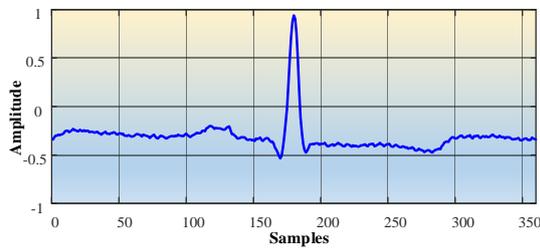


Figure 2. Normal heartbeat ECG signal example.

PVC is an abnormal condition that the heartbeat is initiated by ventricular Purkinje fibers rather than by the sinoatrial node, which is the normal heartbeat initiator. As a result, extra contractions occur, and the regular heart rhythm breaks down. An illustration of the PVC signal is given in Figure 3.

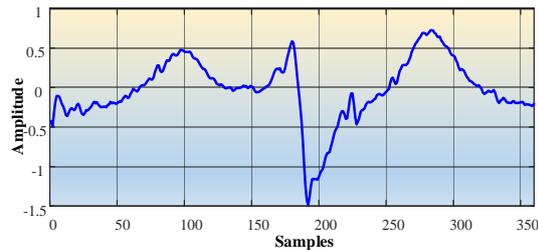


Figure 3. PVC signal form example.

An example ECG signal for paced, another abnormal heartbeat, is shown in Figure 4.

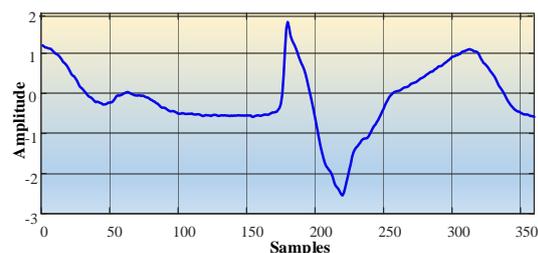


Figure 4. Paced arrhythmia signal example.

The Bundle branch block is a delay in the way of electrical impulses which are ejected to provide a heartbeat. This delay can occur in the right or left ventricles of the heart. If this delay happens in the right ventricles, the RBBB heartbeat shown in Figure 5 occurs, and if this delay happens in the left ventricle, the LBBB heartbeat shown in Figure 6 occurs.

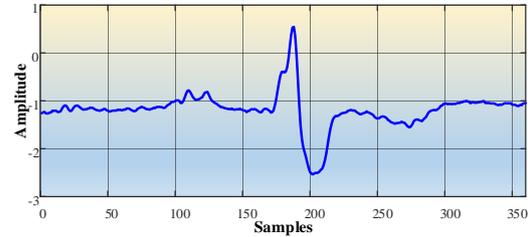


Figure 5. RBBB arrhythmia signal example.

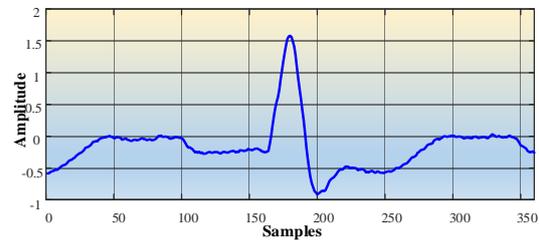


Figure 6. LBBB arrhythmia signal example.

4. Experimental results

In order to determine the optimal features for ECG signals, five different heartbeat classes were selected from the MIT-BIH arrhythmia database. The 50% of the data were used in the training phase while the rest were used for testing. The numerical distributions of data classes are given in Table 1.

Table 1. Number of used ECG signals.

Classes	Train Data	Test Data	Total
Normal	2500	2500	5000
Paced	950	950	1900
PVC	250	250	500
RBBB	2250	2250	4500
LBBB	950	950	1900

Statistical methods have been used on 6-levels Db6 wavelet transform coefficients to generate 28-dimensional feature vectors of ECG signals. The properties and definitions in feature vectors are given in Table 2.

Feature vectors that have 28-dimensional for each heartbeat signal are available as input data. PSO, genetic search, best first, greedy stepwise and multi objective evolutionary algorithms are used for feature optimization. The optimal features obtained after applying these methods to the feature vectors are given in Table 3.

PSO, best first and greedy stepwise methods were

determined 17 feature for this signals. The lowest number of features is determined by the genetic search method while the multi objective evolutionary search algorithm determines nine features. We have tested the classification accuracy with feature vectors which are optimized for evaluating the performance of the

optimization algorithms. For this purpose, LS-SVM and RF classifiers were used in the classification of ECG signals. The recognition accuracy is determined by applying the feature vectors obtained from the optimization algorithms to these classifier inputs.

Table 2. Representation of the features.

Number	Feature	Number	Feature
1...6	E (D ₁ ...D ₆)	7	E (A ₆)
8...13	μ (D ₁ ...D ₆)	14	μ (A ₆)
15...20	σ (D ₁ ...D ₆)	21	σ (A ₆)
22...27	ρ (D ₁ ...D ₆)	28	ρ (A ₆)

E: Energy, μ : Mean, σ : Standard deviation, ρ : Norm entropy, D: Detail band, A: Approximation band.

Table 3. Optimized feature vectors and selected features.

Method	Optimized Feature Vectors	Vector Size
Particle Swarm Optimization	3-5-8-9-10-11-12-13-15-16-18-23-24-25-26-27-29	17
Best First and Greedy Stepwise	2-3-6-9-10-11-12-13-15-17-18-23-24-25-26-27-29	17
Genetic Search	13-15-21-23-25-26-27-28	8
Multi Objective Evolutionary Search	2-3-4-6-10-23-25-26-28	9

Table 4. Performances of the classifiers on the feature vectors.

Feature Vector	Number of Features	Classifiers	
		LS-SVM	RF
All Features	28	97.30%	98.82%
Particle Swarm Optimization	17	97.34%	98.82%
Best First and Greedy Stepwise	17	97.00%	96.79%
Genetic Search	8	88.52%	98.95%
Multi Objective Evolutionary Search	9	80.84%	97.39%

As seen in Table 4, the best performance regarding both feature size and classification performance has been obtained by the genetic search method. RF classifier provides 98.5% performance on ECG datasets having eight features and being optimized by using genetic search. The RF classifier achieved 98.82% performance when the 28-dimensional feature vector containing all the features was given as input. As a result, both feature size has been drastically reduced, and the performance has increased. The PSO algorithm increased the performance from 97.30% to 97.34% with the LS-SVM classifier and reduced 11 features from the feature vector. Other optimization techniques have reduced feature size, but at the same time, they also reduced the performance.

5. Conclusion

In this study, feature vector optimization methods were analyzed in the classification of ECG signals. Five different heartbeat classes from the MIT-BIH arrhythmia dataset were used in experimental studies. Feature vectors containing distinctive features of signals are obtained by using wavelet transform and statistical methods on heartbeat signals. Various optimization algorithms have been used to optimize the 28-dimensional feature vector. Feature vectors obtained from these optimization algorithms are given as input to LS-SVM and RF classifiers. The feature vector, which is a total of 28 dimensions, was reduced to 8 dimensions as a result of genetic search optimization algorithm, resulting in 98.95% performance. When 28 features are used, it is seen that this performance is 98.82%. With the genetic search

optimization algorithm, both the feature vector is reduced, and the recognition performance is improved. In addition, the number of feature vectors is reduced by the PSO algorithm, and the recognition performance is preserved. Another important point is the selection of the classifier. The success achieved by the genetic search algorithm with the LS-SVM classifier was as low as 88.52%, but the success rate with the RF classifier increased to 98.95%.

As a result of experimental studies, it has been observed that the feature vectors significantly affect the performance in recognizing ECG signals. Further, it was shown that how the selected classifier can lead to a better performance on optimized feature vectors.

References

- [1] Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3), 230-236.
- [2] Okada, M. (1979). A digital filter for the QRS complex detection. *IEEE Transactions on Biomedical Engineering*, (12), 700-703.
- [3] Afonso, V. X., Tompkins, W. J., Nguyen, T. Q., & Luo, S. (1999). ECG beat detection using filter banks. *IEEE transactions on biomedical engineering*, 46(2), 192-202.
- [4] Li, C., Zheng, C., & Tai, C. (1995). Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on biomedical Engineering*, 42(1), 21-28.
- [5] Rekik, S., & Ellouze, N. (2016). QRS detection combining entropy criterion and wavelet transform. *International Journal of Signal and Imaging Systems Engineering*, 9(4-5), 299-304.
- [6] Rani, R., Chouhan, V. S., & Sinha, H. P. (2015). Automated detection of QRS complex in ECG signal using wavelet transform. *International Journal of Computer Science and Network Security (IJCSNS)*, 15(1), 1.
- [7] Kaur, I., Rajni, R., & Marwaha, A. (2016). ECG Signal Analysis and Arrhythmia Detection using Wavelet Transform. *Journal of The Institution of Engineers (India): Series B*, 97(4), 499-507.
- [8] Chen, T. H., Zheng, Y., Han, L. Q., Guo, P. Y., & He, X. Y. (2008). The Sorting Method of ECG Signals Based on Neural Network. In *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on* (pp. 543-546). *IEEE*.
- [9] Dokur, Z., Ölmez, T., Yazgan, E., & Ersoy, O. K. (1997). Detection of ECG waveforms by neural networks. *Medical engineering & physics*, 19(8), 738-741.
- [10] Coast, D. A., Stern, R. M., Cano, G. G., & Brillner, S. A. (1990). An approach to cardiac arrhythmia analysis using hidden Markov models. *IEEE Transactions on biomedical Engineering*, 37(9), 826-836.
- [11] Jain, S., Kumar, A., & Bajaj, V. (2016). Technique for QRS complex detection using particle swarm optimisation. *IET Science, Measurement & Technology*, 10(6), 626-636.
- [12] Thomas, M., Das, M. K., & Ari, S. (2015). Automatic ECG arrhythmia classification using dual tree complex wavelet based features. *AEU-International Journal of Electronics and Communications*, 69(4), 715-721..
- [13] Inan, O. T., Giovangrandi, L., & Kovacs, G. T. (2006). Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Transactions on Biomedical Engineering*, 53(12), 2507-2515.
- [14] Sahoo, S., Kanungo, B., Behera, S., & Sabut, S. (2017). Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities. *Measurement*, 108, 55-66.
- [15] Martis, R. J., Acharya, U. R., & Min, L. C. (2013). ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control*, 8(5), 437-448.
- [16] Ince, T., Kiranyaz, S., & Gabbouj, M. (2009). A generic and robust system for automated patient-specific classification of ECG signals. *IEEE Transactions on Biomedical Engineering*, 56(5), 1415-1426.
- [17] Martis, R. J., Acharya, U. R., Mandana, K. M., Ray, A. K., & Chakraborty, C. (2013). Cardiac decision making using higher order spectra. *Biomedical Signal Processing and Control*, 8(2), 193-203.
- [18] Tadejko, P., & Rakowski, W. (2007). Mathematical morphology based ECG feature extraction for the purpose of heartbeat classification. In *Computer Information Systems and Industrial Management Applications, 6th International Conference on* (pp. 322-327). *IEEE*.
- [19] Kim, J., Shin, H., Lee, Y., & Lee, M. (2007). Algorithm for classifying arrhythmia using Extreme Learning Machine and principal component analysis. In *Engineering in Medicine and Biology Society, EMBS 2007. 29th Annual International Conference of the IEEE* (pp. 3257-3260). *IEEE*.
- [20] Martis, R. J., Acharya, U. R., Mandana, K. M., Ray, A. K., & Chakraborty, C. (2012). Application of principal component analysis to ECG signals for automated diagnosis of cardiac health. *Expert Systems with Applications*, 39(14), 11792-11800.
- [21] Mehta, S. S., & Lingayat, N. S. (2008). Development of SVM based ECG Pattern Recognition Technique. *IETE Journal of Research*, 54(1), 5-11.
- [22] Raman, P., & Ghosh, S. (2016). Classification of Heart Diseases based on ECG analysis using FCM and SVM Methods. *International Journal of*

- Engineering Science*, 6739.
- [23] Ceylan, R., & Özbay, Y. (2007). Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. *Expert Systems with Applications*, 33(2), 286-295.
- [24] Shadmand, S., & Mashoufi, B. (2016). A new personalized ECG signal classification algorithm using block-based neural network and particle swarm optimization. *Biomedical Signal Processing and Control*, 25, 12-23.
- [25] Güler, İ., & Übeyli, E. D. (2005). ECG beat classifier designed by combined neural network model. *Pattern recognition*, 38(2), 199-208.
- [26] Moraglio, A., Di Chio, C., & Poli, R. (2007). Geometric particle swarm optimisation. In *European conference on genetic programming* (pp. 125-136). Springer, Berlin, Heidelberg.
- [27] Gutlein, M., Frank, E., Hall, M., & Karwath, A. (2009). Large-scale attribute selection using wrappers. In *Computational Intelligence and Data Mining, IEEE Symposium on* (pp. 332-339). IEEE.
- [28] Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine learning*, 3(2), 95-99.
- [29] Jiménez, F., Sánchez, G., García, J. M., Sciavicco, G., & Miralles, L. (2017). Multi-objective evolutionary feature selection for online sales forecasting. *Neurocomputing*, 234, 75-92.
- [30] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [31] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300.
- [32] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182..
- [33] Martín-Smith, P., Ortega, J., Asensio-Cubero, J., Gan, J. Q., & Ortiz, A. (2017). A supervised filter method for multi-objective feature selection in EEG classification based on multi-resolution analysis for BCI. *Neurocomputing*, 250, 45-56.
- [34] Mark, R. Moody, G. (1997). MIT-BIH Arrhythmia Database, <http://ecg.mit.edu/dbinfo.html>
- Özal Yıldırım** received BSc degree in Electrical and Computer Education, Fırat University, Turkey, in 2006. He received MSc degree in Computer Engineering Department from Fırat University, in 2010. He was awarded PhD degree from Electrical and Electronics Engineering from Fırat University, in 2015. Dr. Yıldırım is working as assistant professor in Computer Engineering Department at Munzur University, Turkey. His main research interests include signal processing, power quality, artificial intelligence and deep learning.
- Ulas Baran Baloglu** received BSc degree in Computer Engineering from I.D. Bilkent University, Ankara, Turkey and the MSc degree in Computer Engineering from Fırat University, Elazığ, Turkey. He was awarded PhD degree from Electrical and Electronics Engineering from Fırat University in 2017. He is working as an assistant professor at the department of Computer Engineering, Munzur University, Turkey. His research interests include multi-agent systems, smart grids, data structures and optimization algorithms. He is IEEE volunteer member.

