Classification of ECG Signals of Normal and Abnormal Subjects Using Common Spatial Pattern

Latifah Aljafar[1], Turky N. Alotaiby[2], Rand R. Al-Yami[1], Saleh A. Alshebeili[3], Jalila Zouhair[1]
[1]Pince Sultan University, Kingdom of Saudi Arabia
[2]KACST, Kingdom of Saudi Arabia.
[3]KACST-TIC in Radio Frequency and Photonics for the e-Society (RFTONICS), Dept. of Electrical Engineering, King Saud University, Kingdom of Saudi Arabia.

Abstract—**In this paper, an ECG signal classification method is presented to classify multi-lead ECG signals into normal and abnormal classes using Common Spatial Pattern (CSP) as the feature extraction algorithm. The method consists of two main stages: CSP-based feature extraction and classification. After segmenting the signal into non-overlapping segments, each segment is projected onto a *CSP* projection matrix to extract the training and testing feature vectors. These vectors are used in the classification stage. In this study, three classifiers—linear discriminant analysis (LDA), naïve Bayes (NB), and support vector machine (SVM)—were used. The proposed approach was evaluated using 104 subjects' recordings (52 normal and 52 abnormal) from the Physikalisch-Technische Bundesanstalt (PTB) dataset. The three classifiers achieved accuracy rates of 80.65%, 84%, and 100%, respectively.**

I. Introduction

Cardiovascular disease is the leading cause of death globally. It accounts for 17.3 million deaths/year, and by 2030, this is expected to rise to 23.6 million [1]. Electrocardiogram (ECG) signals are an important source of information to the cardiologist for cardiovascular disease diagnosis. However, analyzing or processing these signal records manually to discover heart diseases is time-consuming and sensitive to human errors. To solve this issue, an automatic diagnostic classification system is required, and many studies have been proposed. Chen et al. employed the sequential probability ratio test (SPRT) algorithm with blanking variability to classify 70 episodes of ventricular fibrillation (VF) and 30 episodes of ventricular tachycardia (VT) obtained from the MIT arrhythmia database [2]. They achieved an accuracy rate of 95% [3]. Christov and Bortolan [4] presented a classification method for classifying normal QRS complexes and premature ventricular contractions (PVC) using a neural network (NN). They extracted 26 parameters; 23 heartbeat interval-related and 3 vectorcardiogram (VCG) parameters. They evaluated their method using 48 ECG recordings from the MIT-BIH arrhythmia database and achieved an accuracy rate of 99%. Bortolan et al. [5] investigated hyperbox classifiers for classifying two types of ECG heartbeats—normal and PVC—using 26 morphological features. They evaluated their method with a set of 22 ECG records in the MIT-BIH ECG arrhythmia database and achieved an accuracy rate of 99%. Arif et al. employed the back-propagation neural network (BPNN) classifier, which aims to classify both healthy and unhealthy subjects with myocardial infarction (MI) using time-domain ECG features. They evaluated their method using a Physikalisch-Technische Bundesanstalt (PTB) dataset and achieved an accuracy rate of 93.7% [6]. Kumar and Kumaraswamy [7] proposed a classification method using the ECG time-domain feature (RR intervals) and the Feed Forward Neural Network (FF-NN) as a classifier. Their method achieved an accuracy rate of 92.8% using the MIT-BIT arrhythmia database. Khazaee [8] proposed a classification method to classify three types of ECG beats (normal, PVC, and others) using morphological amplitudes and heartbeat interval features. He used a support vector machine (SVM) as a classifier and the particle swarm optimization algorithm as an optimizer using a total of 12 records from the MIT-BIH Arrhythmia database. His method achieved an accuracy rate of 97%. Kaveh and Chung [9] developed a classification method to classify ECG signals as healthy and unhealthy using time and frequency domain features. They evaluated their method using 89 records (46 healthy and 43 unhealthy) from the MIMIC II database [10]. Half of these records were randomly selected to train the SVM classifier, and the unused half was used for testing and achieved an accuracy rate of 88.4%. Harkat and Saidi proposed a method for classifying normal and abnormal arrhythmia beats using time-frequency domain features and the radial basis function network (RBF) as a classifier. The RBF is optimized by the cuckoo search algorithm via Levy flight. They tested their method using 300 beats from each class in the training phase and 800 beats from the normal class and 4100 beats from the abnormal class in the testing phase. Their method achieved an accuracy rate of 98.32 % [11]. Heidarnejad et al. [12] presented a classification method for classifying normal and abnormal (MI) ECG signals based on features extracted from the time and frequency domains and a linear classifier. They reported an accuracy rate of 91% using 100 recording (50 normal and 50 abnormal) from the PTB dataset. Keshtkar et al. [13] employed a probabilistic neural network to classify the ECG signals as normal and abnormal (MI) using six features extracted from signal-averaged electrocardiograms (SAECGs) and their wavelet coefficient. They achieved 93.0% sensitivity, 86.0% specificity, and 89.5% accuracy using the PTB dataset [14]. Bhaskar proposed a method to classify ECG signals as normal and abnormal (MI) using time-frequency features and two classifiers: SVM and a neural network. Their method achieved an accuracy rate of 91% with SVM using 82 healthy controls and 367 myocardial infarction subjects from the PTB dataset [15].

978-1-5090-5306-3/16/$31.00 ©2016 IEEE

In this paper, we propose a classification approach to classify multi-lead ECG signals into normal and abnormal classes based on features extracted using the Common Spatial Pattern (CSP) algorithm. This approach is composed of two main stages: CSP-based feature extraction and classification. The results show that the proposed method with the SVM classifier could be of potential value for cardiologists in the early diagnosis stage of cardiovascular disease. The remainder of the paper is organized as follows. Section II explains the common spatial pattern. Section III presents the methodology. Experimental results are presented in Section IV. Finally, Section V gives concluding remarks.

II. COMMON SPATIAL PATTERN

CSP is a statistical method that has been proven to be an effective feature extraction method [16–19] to discriminate between two classes. The aim of using CSP in this work is to distinguish between normal and abnormal ECG signals. The objective of CSP is to construct a project matrix W that minimizes the variance for one class and maximizes it for another class. The mathematical formulation of CSP is as follows:

1. The normalized covariance matrix for each segment $S \in R^{NxL}$ is calculated.

$$C = \frac{SS^T}{trace(SS^T)} \quad (1)$$

   Where N is the number of leads, L is the number of samples, and T is the transpose.

2. An averaging process is performed on the covariance matrices of each class to find two discriminated covariance matrices, C_1 (normal ECG) and C_2 (abnormal ECG), and obtain composite covariance matrix C_c.

$$C_c = C_1 + C_2 \quad (2)$$

3. Singular value decomposition (SVD) is performed on the matrix C_c to obtain the matrix of Eigenvalues ψ and the normalized Eigenvectors matrix F_c as follows:

$$C_c = F_c \psi F_c^T \quad (3)$$

4. A whitening transform is performed on both covariance matrices C_1 and C_2 with a whitening matrix given by [16,17]:

$$P = \psi^{-1} F_c^T \quad (4)$$

   Such that:

$$D_1 = PC_cP^T \text{ and } D_2 = PC_cP^T \quad (5)$$

   D_1 and D_2 share common eigenvectors. Hence, the sum of the corresponding Eigenvalues of the two matrices is always 1.

5. D_1 and D_2 have the decomposition:

$$D_1 = U\Lambda_1U^T \text{ and } D_2 = U\Lambda_2U^T \quad (6)$$

with $\Lambda_1 + \Lambda_2 = I$, where I is the identity matrix. U and Λ represent the matrix of Eigenvectors and the diagonal matrix of Eigenvalues, respectively.

The Eigenvalues are then sorted in descending order; hence, the projection matrix is formulated as $W = U^T P$ of the size NxN. Each row of W represents a spatial filter, and each column of W^{-1} represents a spatial pattern.

III. METHODOLOGY

The proposed ECG signal classification is composed of two main stages: CSP-based feature extraction and classification.

A. CSP-based feature extraction

The feature extraction stage is the most important stage in the ECG signal classification process. In this step, the multi-lead ECG signals are de-trended by removing the mean value from each lead. Then, for feature extraction with the CSP algorithm, the multi-lead ECG signals need to be segmented into non-overlapping segments. A moving window of three seconds is used for signal framing.

For extracting the CSP features, the transpose of each segment of 15x3000 in size (number of leads x number of samples) is projected onto the projection matrix W of 15x15 in size. The features are extracted by taking the log of the variance of the projected matrix.

B. Classification

The classification stage consists of two phases: training and testing. In the training phase, the classifier(s) are trained on the feature vectors of both normal and abnormal ECG signals. In the testing phase, the classifier(s) are used to classify a new feature vector as a normal or abnormal subject.

C. Data

The dataset used in this study is from a publically available database, PTB; the National Metrology Institute of Germany provided this combination of digitized ECGs for research to the visitors of PhysioNet. It contains normal and abnormal subjects' recordings. All recordings have 15 leads. All signals are sampled at 1000 samples/second with 16-bit resolution. In this research, 52 healthy (normal) subjects, 39 myocardial infarction (MI) patients, and 13 dysrhythmia (DY) patients were recruited. Table 2 represents a summary of the used data.

TABLE I: NUMBER OF TRAINING AND TESTING SEGMENTS

Class	Diagnosis	Training segments	Testing segments	Number of subjects
1	Abnormal	919	920	52
0	Healthy controls (normal)	1017	1018	52
Total		1936	1938	104

IV. EXPERIMENTAL RESULTS

The performance of the proposed method was evaluated using the most widely used metrics: sensitivity (Sen), specificity (Spe), and accuracy (Acc), which are defined as follows:

$$Sen = TP/(TP+FN) \quad (7)$$

$$Spe = TN/(TN+FP) \quad (8)$$

$$Acc = (TP+TN)/(TP+TN+FN+FP) \quad (9)$$

where TP is the number of true positive cases, FN is the number of false negative cases, TN is the number of true negative cases, and FP is the number of false positive cases.

The data (Table I) are divided into training and testing segments with 1936 segments for training and 1938 for testing. In the classification stage, three classifiers were used: linear discriminant analysis (LDA), naïve Bayes (NB), and SVM. Table II shows the test data results for normal and abnormal classes. As can be seen, the proposed method can achieve an accuracy of 80.65% with LDA, 84% with NB, and 100% with SVM. SVM's results demonstrate the effectiveness of the proposed ECG classification algorithm in discriminating between normal and abnormal ECG segments.

TABLE II : SUMMARY OF THE RESULTS

Classifier	Sen.	Spe.	Acc.
LDA	80.94	80.33	80.65
NB	85.17	82.72	84
SVM	100	100	100

V. CONCLUSION

In this paper, we present an automated method for the classification of multi-lead ECG signals as either normal or abnormal using CSP as the feature extractor. The proposed approach consists of two main stages: CSP-based feature extraction and classification. After de-trending and segmenting the ECG signal, each segment is projected onto a projection matrix to obtain the feature vectors. Three classifiers were employed in the classification stage: LDA, NB, and SVM. This method achieved an accuracy rate of 100% using SVM. These results indicate that the proposed approach has the potential to help cardiologists analyze ECG signals.

REFERENCES

[1] https://www.heart.org/idc/groups/ahamah-public/ @wcm/ @sop/@smd/ documents /downloadable/ucm_470704.pdf [Accessed: 4- Oct- 2016].

[2] https://www.physionet.org/physiobank/database/mitdb/. [Accessed: 20- Jun- 2016].

[3] Szi-Wen Chen, P. Clarkson and Qi Fan, "A robust sequential detection algorithm for cardiac arrhythmia classification", IEEE Transactions on Biomedical Engineering, vol. 43, no. 11, pp. 1120-1124, 1996.

[4] I. Christov and G. Bortolan, "Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks", Physiol. Meas., vol. 25, no. 5, pp. 1281-1290, 2004.

[5] G. Bortolan and W. Pedrycz, "Hyperbox classifiers for arrhythmia classification", Kybernetes, vol. 36, no. 34, pp. 531-547, 2007.

[6] M. Arif, I. A. Malagore and F. A. Afsar, "Automatic Detection and Localization of Myocardial Infarction Using Back Propagation Neural Networks," 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), Chengdu,pp. 1-4 , 2010 .

[7] R.G. Kumar, Y.S. Kumaraswamy, "Investigation and classification of ECG beat using input output additional weighted feed forward neural network", in: International Conference on Signal Processing, Image Processing & Pattern Recognition (ICSIPR), pp. 200–205, 2013.

[8] A. Khazaee, "Heart beat classification using particle swarmOptimization", Int. J. Intell. Syst. Appl. 5 (6) (2013) 25.

[9] A. Kaveh and W. Chung, "Automated classification of coronary atherosclerosis using single lead ECG," Wireless Sensor (ICWISE), 2013 IEEE Conference on, Kuching, pp. 108-113 ,2013.

[10] 1 AL. Goldberger, LAN. Amaral, L. Glass, JM. Hausdorff, PCh. Ivanov, RG. Mark, JE. Mietus, GB. Moody, CK Peng, HE. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220.

[11] A. Harkat, R. Benzid and L. Saidi, "Features extraction and classification of ECG beats using CWT combined to RBF neural network optimized by cuckoo search via levy flight," 2015 4th International Conference on Electrical Engineering (ICEE), Boumerdes, pp. 1-4 ,2015.

[12] A. Heidarnejad, S. Setarehdan and V. Nafisi, "Detection of myocardial infarction using high-frequency QRS components", 2011 19th Iranian Conference on Electrical Engineering, pp. 1-5, 2011.

[13] Ahmad Keshtkar, Hadi Seyedarabi, Peyman Sheikhzadeh, and Seyed Hossein Rasta, "Discriminant Analysis Between Myocardial Infarction Patients and Healthy Subjects Using Wavelet Transformed Signal Averaged Electrocardiogram and Probabilistic Neural Network", Journal of Medical Signals and Sensors, vol.3, no. 4, pp. 225–230, 2013.

[14] "The PTB Diagnostic ECG Database", Physionet.org, 2016. [Online]. Available: https://www.physionet.org/physiobank/database/ptbdb/. [Accessed: 20- Jun- 2016].

[15] N. Bhaskar, "Performance Analysis of Support Vector Machine and Neural Networks in Detection of Myocardial Infarction", Procedia Computer Science, vol. 46, pp. 20-30, 2015.

[16] Z. J. Koles, M. S. Lazar, S. Z. Zhou, "Spatial Patterns Underlying Population Differences in the Background EEG," Brain Topography, Vol. 2, No. 4, pp. 275-284, 1990.

[17] Z. J. Koles, J. C. Lind, P. Flor-Henry, "Spatial Patterns in the Background EEG Underlying Mental Disease in Man," Electroencephalograph and Clinical Neurophysiology, Vol. 91, No. 5, pp. 319-28, 1994.

[18] T. Alotaiby, F. El-Samie, S. Alshebeili, K. Aljibreen and E. Alkhanen, "Seizure detection with common spatial pattern and Support Vector Machines", International Conference on Information and Communication Technology Research (ICTRC), pp. 152 - 155, 2015.

[19] S. Wu, C. Wu, N. Pal, C. Chen, S. Chen, "Common spatial pattern and linear discriminant analysis for motor imagery classification", IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB) , pp.146-151, 2013.