2018 International Conference on Big Data and Artificial Intelligence

ECG Signal Classification with Deep Learning for Heart Disease Identification

Wenbo Zhang[§] Limin Yu[*] Lishan Ye[†] Weifen Zhuang[‡] Fei Ma[§]
[*]Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China.
[†]Information Department, Zhongshan Hospital of Xiamen University, Xiamen, 361004, China.
[‡]School of Management, Xiamen University, Xiamen, 361005, China.
[§]Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
Telephone: (86) 512 8816 1633, Email: fei.ma@xjtlu.edu.cn

***Abstract*—Electrocardiogram (ECG) signal is widely used in medical diagnosis of heart diseases. Automatic extraction of relevant and reliable information from ECG signals has not been an easy task for computerized system. This study proposes to use 12-layer 1-d CNN to classify 1 lead individual heartbeat signal into five classes of heart diseases. The proposed method was tested on MIT/BIH arrhythmia database and results were measured using positive predictive value, sensitivity and F1 score. Our proposed method obtained a positive predictive value of 0.977, sensitivity of 0.976, and F1 score of 0.976. Comparing with the results obtained by other four methods on the same database, our method was found superior on all three measures.**

***Keywords*—Electrocardiogram, 1D Convolutional neural networks (CNNs), ECG classification, Heart diseases diagnosis**

I. Introduction

Although electrocardiogram (ECG) signal is widely used in medical diagnosis of heart diseases, there are many challenges in extracting relevant and reliable information from ECG signals automatically using computer programs. In the cardiac cycle, each heartbeat constitutes distinct electrical depolarization-repolarization patterns. This property could illustrate heart's electrical activities. Particularly, the morphological characteristics of heartbeat is different from person to person. Even for the same subject, there are several time-varying characteristics, such as the shapes of QRS complex and R-R interval. They would also change under different circumstances[1]. For experienced doctors, any anomaly could be easily detected by heart pulse or transformation in the morphological pattern. Nevertheless, it is not an easy task for automatic computerized system, since there are external noise and imbalanced classes in the data set.

In the last decade, in order to decrease labor cost and the objectivity of detection, some machine learning algorithms have been used in this area to detect and classify arrhythmia. The usual process of these approaches are based on three main steps: preprocessing, feature extraction, and classification. Firstly, preprocessing methods are applied to de-noise the ECG signals. Yan Lu et al. conducted experiments on both real and synthetic ECG signal using Empirical Mode Decomposition [2], which demonstrated that this approach is superior tool for ECG denoising. Inan utilized wavelet transform on ECG data and achieved 96.82% accuracy on test data set [3]. After denoising, the ECG waveforms are extracted by means of segmentation. The waveforms are then used to calculate and generate features. In the meantime, dimension reduction algorithms could be applied to reduce relatively independent features. Chawla combined principal component analysis (PCA), independent component analysis (ICA) into PCA-ICA algorithm, a better method of feature extraction [4]. Kaur and Arora used linear discriminant analysis (LDA) to reduce features and used machine learning methods on remaining features to construct a classification model [5]. They achieved an average classification accuracy of 99.056%. Osowski et al. used support vector machine (SVM) as a recognition system to recognize heartbeat and confirmed its reliability and advantage in the experiment [6]. Coast et al. proposed a new approach to cardiac arrhythmia analysis based on hidden Markov models, which provided improve performance compared with initial results [7]. Kumar et. investigated cardiac arrhythmia in ECG using random forest and reported accurate results [8]. Hu et al. combined a small customized classifier with a global one to form a integrated system on ECG data [9]. Nonnegative matrix factorization based classification method was firstly proposed by Lee and Seung [10] and has also been applied on many classification tasks [11], [12]. Wang [13] used nonnegative matrix factorization method to automatically identify electrocardiosignal and reached an accuracy of 95.64%. However, some above methods do not perform very well when standard tests are used to evaluate their performance. Also, identify and extract useful features is complex and time consuming.

With the aim of solving such drawbacks and deficiencies, a novel deep learning classifier based on multi-layer convolutional neural networks (CNNs) is proposed in this paper. CNNs, a deep feed-forward artificial neural network, was inspired by biological processes and has successfully been applied to analyzing images, especially object recognition and localization. As for using CNNs to deal with signals, Mirowski et al. used convolutional networks for epileptic seizure prediction from EEG signals [14]. They compared

978-1-5386-6136-9/18/$31.00 ©2018 IEEE

CNNs with many traditional algorithms and proved that this could also work on time signal. Kiranyaz et al. trained dedicated simple CNNs for every patient to automatically classify ECG signals [15]. The final results revealed the model was generic because of its simple structure and parameter invariant property. Rajpurkar et al. built a complex deep learning model with convolution layers based on their own data sets [16]. The results of model exceeded the average cardiologist performance in both precision (positive predictive value) and recall rate (sensitivity). Jiang and Kong proposed an evolvable block-based neural networks for ECG heart beat classification [17]. Ince et al. used wavelet transform in the process of extracting features, with principal component analysis to reduce dimension of features, and designed classifier based on feedforward and fully connected artificial neural networks [18].

The rest of this paper is organized as follows. In Section II, the well-known MIT-BIH arrhythmia database used in the experiment, and the preprocessing of data are outlined. The structure of multi-layer 1-D CNNs and other layers are shown in Section III. Then in Section IV, proposed method are evaluated with the test data by standard multi-class metrics,in terms of its performance and robustness. The results are compared with other approaches. Finally, the conclusion is presented in Section V.

II. Data set and Summary Statistics

A. Data preprocessing

The data set we utilized in this study is from MIT/BIH arrhythmia database [19]. The data set contains 48 records studied by the BIH Arrhythmia Laboratory. Each record contains two-channel 30-min duration ECG signal selected from 24-h recordings of each patient. The frequency of ECG signal is 360Hz. The database do not only provides ECG signal but also corresponding timing information and heartbeat class information verified by cardiologist. From MIT/BIH arrhythmia database, 44 records were picked to use. These records contain various categories of heart beat, such as normal beat, ventricular, atrial arrhythmias. Four records containing paced heartbeats were excluded to use in this study.

For this research, we only use lead-1 ECG signal as our raw data. Considering there are external noises among ECG signals, de-noising method was applied to get clean data. The wavelet transform is a de-noising method which uses wavelet function to convolve with the signal. We use this method in this study since it can keep specific detailed time frequency components of ECG signal [20]. Also, wavelet transform could extract wavelet coefficients and reconstruct signal to eliminate baseline drift [21]. As shown in Figure 1, the red signal is raw data and the green one is de-noised signal. Compared with the signal before de-nosing, the processed signal is more smooth and fluctuates around 0-level curve, which indicates that wavelet transform is effective. Therefore, the de-noised signal avoids the interferences of noises. In order to ensure the precision and accuracy of the experiment, the processed data were utilized in this research.

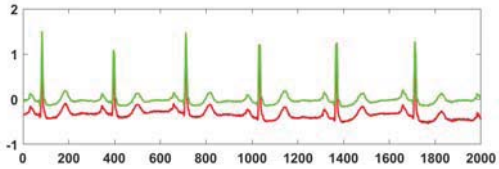

Fig. 1. Raw ECG signal and the de-noised signal using wavelet transform

To get the training data set, we transform continuous ECG signal to individual heartbeat. The width of single beat was approximated to 300 sample data and the beat is centered around R peak [22]. Therefore, we use the R-peak annotation offered by the database to do the transformation. Since these R-peak points are regarded as center point of each heartbeat, for each beat, we cut off continuous signal to obtain 150 sample points before R-peak and 149 after it, meaning that one beat has 300 features. A total of 102548 heartbeats are obtained. Table I shows the number of samples in 10 categories of heart diseases as recommended by AAMI [23]. Under each category, there are different classes of heart diseases. The number of samples of each class is different. We picked five classes from all classes that have largest number of samples for the purpose of more efficient training. Corresponding beats are extracted from all heartbeats and are split into two parts after shuffled. One is used as training set with 70% data, and the other is held as testing set. The number of samples of the selected sets are listed in Table II.

TABLE I
THE NUMBER OF SAMPLES IN EACH CLASS

Normal Rhythm	80921
Atrial Rhythm	2745
Ventricular Rhythm	8538
Arrhythmia	335
Cannot analysis	7187
Noise	586
block	193
Abnormal A	1241
Abnormal B	982
sum	102548

TABLE II
THE NUMBER OF SAMPLES IN EACH SELECTED CLASS

Classes	training data	test data
Normal Beat	48755	20895
Left Bundle branch block beat	2576	1104
Right bundle branch block beat	5092	2182
Premature ventricular contraction beat	4867	2086
Atrial premature beat	1782	764
total number	63072	27031

B. Rhythm Classes

As mentioned before, we picked five most representative classes: normal beat, left buddle branch block beat (LBBB), right buddle branch block beat (RBBB), premature ventricular contraction (PVC) and atrial premature beat (APB). The main reasons we choose them is that these five classes account for most samples among 102548 heart beats. There are two advantages of this choice. Firstly, it is more convenient and effective to train CNNs with large data sets rather than small ones. Hence classes with most samples were picked to ensure the CNNs could be trained better. Secondly, it can mitigate the influence of imbalanced classes. Since if the numbers of samples of each class have a huge difference, it would cause classification model to be totally useless. In order to address this problem, in this study we neglect the classes with a small size and pick the large ones.

III. Model

A. Overview of CNN

The traditional CNNs are inspired by feedforward artificial neural network that presents a structure of mammalian visual cortex. They have widely become the standard model in image classification and video detection system. There are three important advantages of CNNs compared with tradition machine learning algorithm: sparse interactions, parameter sharing and equivariant representations [17]. The input for each neuron in 2-D CNNS is 2-D matrixes, which is not suitable in our task, since the ECG data is only one dimension. Therefore, we replace the tradition 2-D CNNs with 1-D so our processed ECG can be used as input.

B. Problem formulation

For the ECG arrhythmia detection, the input data is a sequence of raw ECG signal denoised by wavelet transform, and output is a sequence of probability which corresponds to the label. For the single sample in the training set, the cost function we chose is cross-entropy objective function:

$$L(\{x,y\}_1^N) = -\sum_n \sum_i y_i^{(n)} * \log(\hat{y}_i^{(n)}) \tag{1}$$

where $\{x,y\}_1^N$ are the training data with corresponding label, $y_i^{(n)}$ represents whether n-th sample belongs to class i, and the value would be 1 if it is true otherwise be 0. $\hat{y}_i^{(n)}$ is the probability that the network assigns the n-th sample to i-th class,

C. Model Architecture

As mentioned before, the model we designed is a multi-layer 1-d CNN with pooling layers, dropout layers and fully connected layers, which is presented in Figure 2. The network takes a time series of ECG signal as input data, and outputs a sequence of label prediction. The network consists of 12 layers including 4 1-D convolution layers. Before data are fed into input layer, in order to increase the speed of training, it is efficient to use standardization. Here we standardized the data using Z-score:

$$x^* = \frac{x-\mu}{\sigma} \tag{2}$$

where x is the initial data, μ and σ are the mean and standard deviation of the data respectively. We used standardized data in the following experiment.

Each convolution layer in our network has 256 filters of size 3*3, since filters with small size could decrease the computation to train the model faster. For each convolution layer, we use rectified linear units(relu) rather than sigmoid as its activation function, since compared with sigmoid, rectified linear units solve the problem of gradient vanishing and decrease the efforts of computing [24]. In the beginning of the network, we piled two convolution layers. The reason why we design it in this way is that this structure has stronger ability of learning features. For example, compared with one layer, this structure incorporates two non-linear relu activation functions rather than single one, meaning that the decision function could be more discriminative [25]. After convolution layer, a pooling layer follows it, since it makes the representation become approximately invariant to small translations of the input [24]. Also, pooling layers could decrease the dimension of data, which is useful to avoid over-fitting and decrease the amount of computing, meaning that it is beneficial to train the model faster. In these layers, we use filters of size 2*2 [26]. Dropout layer is used to randomly set some dimensions of input vector to be zero with a probability. It does not have any trainable parameters, meaning that nothing in this layer would update during training. This kind of layer can mitigate over-fitting to a large extent on many benchmark tests [27]. The final fully connected layer with 120 neurons followed by one output layer produces a distribution over the 5 output labels. To accomodate multi-class classification, the activation function we use in the output layer is softmax:

$$y(z) = \frac{e_i^z}{\sum_j e_j^z} \tag{3}$$

The figure of model is shown in Fig. 2. During the training process, Adam optimizers [28] was used for the parameters. The best model with best performance on the validation set during the optimization process was kept.

IV. Experiments

In the experiment, the preprocessed training data set was fed into CNNs model. After training, the model was used on test data set to obtain test results. After testing, a series metric were used to evaluate the performance of the model.

A. Evaluation metrics

In this study, annotations provided by MIT-BIH are used to evaluate the results obtained by proposed model and the

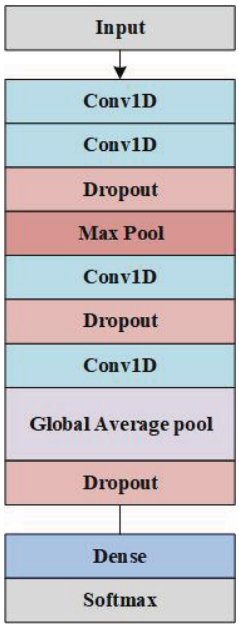

Fig. 2. The structure of the model

performance are captured by using three metrics, positive predictive value (PPV), Sensitivity and F1 score.

In the following formulas, true positive (TP) means the number of items correctly labeled when they belong to the positive class. False positive (FP) refers to items belonging to the positive class that were incorrectly labeled as negative. True negative(TN) and false negative (FN) are items belonging to negative class that were classified as negative and positive respectively. The definitions of three metrics are as follows: Positive predictive value (PPV): the rate of correctly identified positives in all identified positives.

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

Sensitivity: the proportion of positives that are correctly identified

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

F1 Score: A measure of test accuracy. It is the harmonic mean of PPV and sensitivity.

$$F1\ Score = \frac{2 * PPV * Sensitivity}{PPV + Sensitivity} \tag{6}$$

These three metrics are commonly used in medical test area. As there is a huge difference in the number of heartbeats in distinct classes, comparison of accuracy with positive predictive value, sensitivity and F1 score are more relevant performance criteria for arrhythmia detection.

B. Experimental Results

Table II presents the number of samples from each class in training and test data set. The training data set consists of 63072 heartbeats and testing data set includes 27031 heartbeats. Table III shows the comparison matrix of classification results on testing data. The value of positive predictive value, Sensitivity and F1 score are 0.977, 0.976 and 0.976 respectively.

To better check the performance of the proposed model, the results were compared with results obtained by four different classification algorithms, Kiranyaz [15], Hu [9], Jiang [17] and Ince [18], all of which used the same ECG database and evaluation metrics. From the results shown in Table IV, the model outperforms performances of other four methods in all three metrics: positive predictive value, sensitivity and F1 score. However, it is worth to note that this study used the Lead 1 ECG signal, while the comparison methods could use Lead 2 ECG signal or both Lead 1 and Lead 2 ECG signal. Also this study only focused on the classification of five classes of diseases, while the comparison methods could focus on different classes of heart diseases.

TABLE III
THE HEARTBEAT CLASSIFICATION RESULTS ON THE TESTING DATA SET

category	Ppv	Sen	F1 Score
Normal Beat	0.986	0.987	0.986
Left Bundle branch block beat	0.978	0.991	0.984
Right bundle branch block beat	0.884	0.960	0.920
Premature ventricular contraction beat	0.984	0.936	0.959
Atrial premature beat	0.962	0.819	0.885
Average/total number	0.977	0.976	0.976

TABLE IV
PERFORMANCE OF THE PROPOSED METHOD AND OTHER FOUR METHODS FROM THE LITERATURE

Methods	Ppv	Sen	F1 Score
Kiranyaz	0.962	0.959	0.960
Hu	0.758	0.789	0.773
Jiang	0.958	0.943	0.950
Ince	0.922	0.903	0.912
Proposed	0.977	0.976	0.976

V. CONCLUSION

In this study, an ECG heartbeat classifier based on 1-D CNNs was proposed. The model receives one dimension signal and outputs classification results. By using 1-D CNNs, extracting features by automatic filters becomes more efficient. Compared with arrhythmia detection systems based on traditional machine learning methods, the proposed model does not rely on specific image features to realize classification results. In our model, wavelet-based pre-processing improved the system performance by effective de-noising of the raw data. The simulation results demonstrated the performance gain over four comparing methods proposed in recent literature for three performance metrics. The results show that CNN is very useful to deal with local features [27]. Local

feature captures spatial information among neighborhood of pixels, which describes local characters of the input data. If the position of the data points were changed in the input data, the information extracted by the model will be different. In our task, the signals extracted from ECG heartbeats are continuous signals, which are time sequence signals and are all local features. 1-D CNN can accurately capture the characteristics contained among the local signals. The results also suggest that the proposed automatic classification method can be used in heart diseases detection, hence can potentially save a lot of doctor's work.

In this study the proposed model was only tested on the MIT-BIH data set. Only five classes which contain relatively more testing samples were used for experiment. Other classes containing less data, such as Fusion of Ventricular, Ventricular Fibrillation or a number of other heart diseases which do not necessarily exhibit as arrhythmias, were not tested in this study. A data set with more classes and more samples will be able to better evaluate the proposed model and better evaluate the proposed model on different classes of heart diseases. Unfortunately such a database is not available in this study. Also only one lead raw data was used in this study. Some types of signal may be difficult or even impossible to detect on a single-lead but can be seen on multiple-lead ECG. In our future work, we will apply our model to multiple-lead ECG signals.

ACKNOWLEDGMENT

This study is supported by the National Natural Science Foundation of China (Grant No. 61501380, 71672160).

REFERENCES

[1] R. Hoekema, G. J. H. Uijen, and A. V. Oosterom, "Geometrical aspects of the inter individual variability of multilead ecg recordings," *IEEE Trans. Biomed. Eng*, vol. 48, pp. 551–559, May 1999.

[2] Y. Lu, J. Yan, and Y. Yam, "Model-based ecg denoising using empirical mode decomposition," pp. 191–196, Nov 2009.

[3] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Trans. Biomed. Eng*, vol. 53, pp. 2507–2515, Dec 2006.

[4] M. P. S. Chawla, "Pca and ica processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison," *Applied Soft Computing*, vol. 11, pp. 2216–2226, 2011.

[5] M. Kaur and A. S. Arora, "Classification of ecg signals using lda with factor analysis method as feature reduction technique," *Journal of Medical Engineering &Technology*, vol. 36, pp. 411–420, 2012.

[6] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng*, vol. 51, pp. 582–589, April 2004.

[7] D. A. Coast, R. M. Stern, G. G. Cano, and S. A. Briller, "An approach to cardiac arrhythmia analysis using hidden markov models," *IEEE Trans. Biomed. Eng*, vol. 37, pp. 826–836, Sept 1990.

[8] R. G. Kumar and Y. S. Kumaraswamy, "Investigating cardiac arrhythmia in ecg using random forest classification," *International Journal of Computer Applications*, vol. 37, pp. 31–34, 2012.

[9] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ecg beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng*, vol. 44, pp. 891–900, Sept 1997.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix," *factorization[J]. Nature*, vol. 401, pp. 788–791, 1999.

[11] ——, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[12] Y. Li and A. Ngom, "Classification approach based on non-negative least squares," *[J]. Neurocomputing*, vol. 118, pp. 41–57, 2013.

[13] C. Wang, *Electrocardiogram (ECG) Recognition and Classification based on Non-negative Matrix Factorization and Decision Forest*. Thesis, Xiamen University, 2017.

[14] P. W. Mirowski, D. M. Y. Lecun, and R. Kuzniecky, "Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg," pp. 244–249, Oct 2008.

[15] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Trans. Biomed. Eng*, vol. 63, pp. 664–675, March 2016.

[16] P. Rajpurkar, M. H. A. Y. Hannun, and A. Y. N. C. Bourn, "Cardiologist-level arrhythmia detection with convolutional neural networks," 2017.

[17] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ecg signal classification," *IEEE Trans. Neural Netw*, vol. 18, pp. 1750–1761, Nov 2007.

[18] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of electrocardiogram signals," *IEEE Trans. Biomed. Eng*, vol. 56, pp. 1415–1426, May 2009.

[19] R. Mark and G. Moody, "Mit-bih arrhythmia database directory," 2010. [Online]. Available: http://ecg.mit.edu/dbinfo.html

[20] C. Saritha, V. Sukanya, and Y. N. Murthy, "Ecg signal analysis using wavelet transforms," *Bulg.j.phys.chapter*, vol. 35, pp. 155–163, 2008.

[21] M. A. Tinati and B. Mozaffary, "A wavelet packets approach to electrocardiograph baseline drift cancellation," *International Journal of Biomedical Imaging*, vol. 2006, p. 97157, 2006.

[22] N. V. Thakor and Y. Zhu, "Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection," *IEEE Trans. Biomed. Eng*, vol. 38, pp. 785–794, Aug 1991.

[23] *Recommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms*. Arlington, VA, USA:Association for the. Advancement of Medical Instrumentation, 1987.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[26] T. N. Sainath, A. R. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for lvcsr," pp. 8614–8618, 2013.

[27] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout:a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[28] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *international conference on learning representations*, 2015.