An ECG arrhythmia image classification system based on convolutional neural network

Ximing Yu*
Signal and Information Processing Specialty, Communication University of China, Beijing 10024, China
yxm8989@126.com

Abstract. The automatic classification of ECG images has great significance for doctors to diagnose cardiac diseases. In order to improve the accuracy and efficiency of disease diagnosis, this paper presents an automatic classification system for ECG arrhythmia images. The database of MIT-BIH is processed visually and a waveform detection method is proposed for detecting the QRS waveform. Wavelet transform is used to detect the R-R interval and locate the R point peak. A convolutional neural network (CNN) model was built to train and classify the ECG images. Experimental results show that according to the ANSI/AAMI EC57 evaluation criteria, the accuracy (Acc) rate of ventricular ectopic beat (VEB) can reach 95.9%. The sensitivity (Se) evaluation index is 93.0%. The specificity (Spe) evaluation index is 91.9%. For the supraventricular ectopic beat (SVEB) class, the accuracy rate is 93.2%, the sensitivity evaluation index is 81.3%, and the specificity evaluation index is 90.5%.

1. Introduction
Many methods have been used in the classification of arrhythmia images. Yang and Chen [1] put forward a method of discrete wavelet transform to extract the corresponding characteristics of ECG. They classified two types of heartbeat images which are normal beat and paced beat from the MIT-BIH database. Radial basis neural network (RBF) was used to classify the ECG images. Yang et al. [2] artificially reduced the noise of ECG images by using a main component analysis network for feature extraction. In order to improve the processing speed of the classifier, a linear support vector machine method was applied. In the experiment, five kinds of waveforms in the MIT-BIH image database were identified. They can reach 1532.2s average training time and 97.8% accuracy rate. Osowski et al. [3] used support vector machine as the classifier, and used two different methods of data preprocessing, namely high-order statistics and Hermite function. Through experiments on 12 different arrhythmia beats in MIT-BIH database, classification accuracy could be significantly improved. Polat and Gne [4] improved the original SVM method. They proposed a least square support vector machine (LS-SVM), and applied the data dimensionality reduction algorithm of principal component analysis (PCA) to reduce 279 extracted features to 15. Their method reduces the amount of data and increases the speed of operations. A 1-d convolution neural network (CNN) was proposed by Kiranyas et al. [5]. They used the model in the field of ECG images classification. They put the feature extraction and classification of the two main module integration into a single body of learning. The advantage of the CNN model is that it need not handmade feature extracting and pretreatment. A classification system based on double coupling method of CNN was proposed by Zhai and Tin [6]. They convert electrical signals into double input in the form of a 2-d coupled matrix to CNN classifier, and capture the morphological characteristics of the ECG signal and relevance. Jun and Nguyen [7] proposed a method of 2d
convolutional neural network. They converted each ECG image into 2d gray image and input to the CNN classifier. They used some deep learning techniques such as batch standardization, data enhancement, Xavier initialization. The experiment results of MIT-BIH arrhythmia database show that the speed and the calculation efficiency of the method are improved, effectively reduced the computational complexity, and can be used for real-time heart medical monitoring and anomaly detection.

This paper proposed a system of QRS waveform location and visualization for processing the MIT-BIH arrhythmia database. A process of five different ECG arrhythmia images segmentation is proposed to process the ECG data. Then a CNN model architecture is proposed to classify ECG arrhythmia images.

The remaining parts are arranged as follows. The second part introduces the method of processing MIT-BIH arrhythmia data including the QRS waveform positioning visualization and different types of heartbeat segmentation. The third part introduces the structure of convolutional neural network model. The experimental results and the relevant evaluation analysis are in the fourth part. The last part is a conclusion of this paper.

2. The method of ECG data processing

2.1. MIT-BIH arrhythmia database
The MIT-BIH arrhythmia database [8] is recognized as the world’s first universal standard for evaluating the performance of arrhythmias. The database contains 48 sets of ECG records. Each set of records lasts about 30 minutes. They are from 47 persons (Records 201 and 202 refer to the same person) studied by the BIH arrhythmia laboratory between 1975 and 1979. The sampling frequency of the records is 360 Hz.

2.2. The standard of ANSI/AAMI EC57
American National Standards Institute (ANSI) and the Association for the Advancement of Medical Instrumentation (AAMI) proposed an evaluative standard [9] between 1984 and 1987. The purpose is to facilitate the horizontal comparison among different arrhythmia automatic detection algorithms. The ANSI/AAMI EC57 standard changes the usage mode of MIT-BIH arrhythmia database. Four of the 48 records are from the patients who wear heart pacemaker. We should remove them and only reserve the other 44 records when we do our images classification experiment. The number of MIT-BIH arrhythmia database labels are 23 that including 15 kinds of beat labels and 8 kinds of non-beat labels. We should aim at the 15 kinds of heartbeat when carry out the arrhythmia classification work. According to the ANSI/AAMI EC57 standards, we need to map the 15 kinds of heartbeat into the 5 broad categories. The 5 categories include:

N=a normal beat or a bundle branch block beat.
S=a supraventricular ectopic beat, an atrial or nodal (junctional) premature escape beat, or an aberrated atrial premature beat.
V=a ventricular ectopic beat, a ventricular premature beat, a R-on-T ventricular premature beat, or a ventricular escape beat.
F=a fusion of a ventricular and a normal beat.
Q=a paced beat, a fusion paced beat and a normal beat, or a beat that cannot be classified.

2.3. ECG beat interception and classification
Since the original MIT-BIH arrhythmia data is the continuous data of 30 minutes, it should be processed firstly to input into the classifier more effective. PhysioBank is a project resource supported by the national academy of bioengineering. When we use the database in the PhysioBank, signal data visualization is the first step. WFDB tools for python is a GPL successor to the MIT-BIH database package. The effect of WFDB package is to read physiological signals such as electrocardiograms. The major components of the WFDB Software Package are WFDB library. It mainly applications for
signal processing and automated analysis, and the wave software for viewing, annotation, and interactive analysis of waveform data. The waveform data and corresponding annotation are read by the WFDB package for python. It is also possible to change the length of the waveform by setting the sampling point during the visualization phase of the waveform. Fig.1 shows the original data waveform visualization.

An ECG image contains several bands of P-QRS-T, among which the QRS group and R wave peak contain the vast majority of important information. Therefore, it is essential to locate the QRS waveform and detect the R wave peak. Experiments show that positioning of QRS waveform cannot be conducted by the method of finding the maximum value, but locate the band according to the heartbeat correction process. A wavelet transform method is used to detect and correct the QRS waveform, and the instantaneous heart rate is calculated. A QRS waveform detection project is used to locate the R-waves in the first channel of the ECG and modify them move to the local maximum peak, using the number of beats per unit time of high frequency as the search radius. Instantaneous heart rate is the inverse of the interval between two adjacent periods, the reciprocal of two adjacent R-R interphase of the ECG. The frequency is shown below.

\[
F = \frac{1}{T} = \frac{60}{num/\text{min}}
\]

The parameter T in the formula is the R-R interval. The instantaneous heart rate can be detected if there is a small change in the interval between two heart beats. QRS waveform detection before correction is shown in Fig.2 and QRS waveform detection after correction is shown in Fig.3.

![Fig.1 The original data waveform visualization](image1)

![Fig.2 QRS waveform detection before correction.](image2)
The red line represents the located R peak position, and the pink line represents the calculated instantaneous heart rate value. After the detection of QRS waveform, R-R intervals are calculated to determine an image capture flow path. The statistical results are shown in Table 1. The 262 sampling point occurrence frequency is the highest. R peak position is selected as the center. We take 131 sampling points both to the left and to the right. So the original images are split into quantities of small rectangular windows. One-dimensional array of each time and sampling points are collected and inputted into the network for training. The process of image intercept is shown in Fig. 4. The visualization of experimental results is shown in Fig. 5.

Table 1. RR interval length frequency

| Number of sampling points | 262 | 259 | 260 | 263 | 256 | 258 | 264 |
|---------------------------|-----|-----|-----|-----|-----|-----|-----|
| frequency                 | 823 | 817 | 808 | 806 | 799 | 792 | 787 |

Fig. 3 QRS waveform detection after correction.

Fig. 4 image intercept
3. Architecture of the CNN model

When we build CNN classifier construction, an 8-layer network model is adopted. The first convolutional layer has a convolution kernel of 11*11*3, and each convolution of the original image generates a new feature map. The step size of the convolution kernel is 4 pixels, and the convolution is carried out in the horizontal and vertical directions. These pixel layers also need to be processed by pool operation. The scale of pool operation is set as 3*3 in advance, the step size of operation is 2, and the size of the pooling image is 27*27*96. The generated layer is processed by local response normalization layer and the scale of normalization operation is 5*5 size. The input data in the second layer is the 27*27*96 pixel layer output from the first layer. For convenience of subsequent processing, each pixel layer is filled up with picture element. Each set of pixel data is convoluted by a convolution kernel of 5*5*48. So there are 256 convolution cores in total and 27*27*128 two group of pixel layers. After the pooling layer, the size of the pooling pixel is two groups of 13*13*128 pixel layers. The third layer’s pixel data is convoluted with a kernel of 3*3*128. The convolution operation is carried out in the same way as the first layer. There are 384 convolution cores in total and 13*13*192 sets of pixel layers. The generated layer serve as the input data for the fourth layer. Each set of pixel data is convoluted by a convolution kernel of 3*3*192, and there are 384 convolution cores in total. The input data of the fifth layer is the two sets of 13*13*192 pixel layers of the output of the fourth layer. There are 256 convolutional cores in total and two groups of pixel layers of 13*13*128. The size of the input data in the sixth layer is 6*6*256. A total of 4096 filters of size 6*6*256 convolve the input data through the output operation results of 4096 neurons. Then the 4096 output values of this layer are inputed through ReLU activation function and dropout operation. The sixth layer is called the fully connected layer. The 4096 output data from sixth layer are fully connected with 4096 neurons in seventh layer, and then processed by ReLU and working procedure called dropout to generate 4096 data. The 4096 data input in seventh layer are fully connected with 1000 neurons in eighth layer. The structure of the CNN model is shown in Fig.6.
4. Experimental result

4.1. Evaluation metrics
In the classification problem of machine learning, we mainly use some parameters to evaluate the performance of the proposed classifier. In my experiment, three statistical parameters, namely, classification accuracy (Acc), sensitivity (Se) and specificity (Spe) were calculated for analysis and comparison to evaluate the performance of the classifier. These statistical indices were defined in following equations:

Accuracy(%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%  \quad (2)

Sensitivity(%) = \frac{TP}{TP + FN} \times 100\%  \quad (3)

Specificity(%) = \frac{TN}{FP + TN} \times 100\%  \quad (4)

The parameters in the equations mean:
- TP(True Positive): The number of positive samples correctly classified. The predicted positive samples are actually positive samples.
- FP(False Positive): The number of negative samples that are incorrectly labeled as positive samples.
- TN(True Negative): The number of negative samples correctly classified. The predicted negative samples are actually negative samples.
- FN(False Negative): The number of positive samples that are incorrectly labeled as negative samples.

4.2. Experimental platform and result
The experiment software runs with Keras and Tensorflow. The hardware platform is Intel(R) Core(TM) i7-7500U CPU, the main frequency is 2.90GHZ, and the memory is 8.00GB. The category Q waveform changes irregularly. It is meaningless to classify this kind of waveform in practical application. So the experiment gets rid of this category and simplify the problem as a four classification(N, S, V, F) problem. In the experiment, we selected 8691 images as the training set and 1857 images as the testing set. The batch size is 16. The learning rate is 0.01. The rate of dropout is 0.25. The procedure of dropout can effectively reduce the occurrence of overfitting. Since overfitting tends to occur in the full connection layer, we use the method of dropout to randomly delete some hidden neurons in the network. Some of the features might have relied on a combination of implicit nodes in a fixed relationship, and they effectively prevented some features from being effective in the presence of others. The category S and category V are of great significance in medical practical
applications. We calculate the Acc, Se, and Spe assessment criteria of the two important categories. The confusion matrix of the classification problem is shown in Fig.7. And the calculated evaluation indicators are shown in Table.2.

![Fig.7 the classification confusion matrix](image)

In the matrix, the row represents the true value and the column represents the predicted value. The numbers of ‘0,1,2,3’ refer to the classes of ‘N,S,V,F’. The accuracy rate of ventricular ectopic beat can reach 95.9% and the sensitivity evaluation is 93.0%. For the supraventricular ectopic beat class, the accuracy rate is 93.2% and the sensitivity evaluation is 81.3%. We evaluated and analyzed the two important parameters Acc and Se. We can see that the category VEB has the better consequence. This is because VEB types have more specific image characteristics. Comparatively speaking, the features of SVEB images are not particularly obvious.

Table.2 calculated evaluation indicators

|        | Accuracy(Acc) | Sensitivity(Se) | Specificity(Spe) |
|--------|---------------|----------------|-----------------|
| VEB    | 95.9%         | 93.0%          | 91.9%           |
| SVEB   | 93.2%         | 81.3%          | 90.5%           |

5. Conclusion

In this paper, using the MIT-BIH arrhythmia database, we have proposed a system for the automatic processing of the ECG for the classification of arrhythmia images. The database of MIT-BIH is processed visually and a waveform detection method is proposed for detecting the QRS waveform. A CNN model was built to train and classify the ECG images. Experimental results show that according to the ANSI/AAMI EC57 evaluation criteria, The accuracy rate of ventricular ectopic beat can reach 95.9% and the sensitivity evaluation is 93.0%. For the supraventricular ectopic beat class, the accuracy rate is 93.2% and the sensitivity evaluation is 81.3%.

References

[1] G.Yang and Y.Chen, The Study of Electrocardiograph Based on Radial Basis Function Neural Network, Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 143-145, 2010.
[2] W.Y.Yang, Y.J.Si and D.Wang, Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine, Computers in biology and medicine Vol.101 P22-320010-4825, 2018.
[3] S.Osowski, L.T.Hoai and T.Markiewicz, Support vector machine-based expert system for reliable heartbeat recognition, in IEEE Transactions on Biomedical Engineering, vol. 51, no. 4, pp. 582-589, April 2004.
[4] K.Polat and S.Gne, Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine, Applied
Mathematics and Computation 186(1):898-906, 2007.

[5] S.Kiranyaz, T.Ince and M.Gabbouj, Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks, in IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664-675, March 2016.

[6] X.Zhai and C.Tin, Automated ECG Classification Using Dual Heartbeat Coupling Based on Convolutional Neural Network, in IEEE Access, vol. 6, pp. 27465-27472, 2018.

[7] T.J.Jun and H.M.Nguyen, ECG arrhythmia classification using a 2-D convolutional neural network, arXiv:1804.06812v1 [cs.CV] 18 Apr 2018.

[8] B.Moody and G.Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng.Med.Biol.Mag.20,45-50, 2001.

[9] American National Standard for Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. AAMI/ANSI Standard EC57: American National Standards Institute and the Association for the Advancement of Medical Instrumentation, 1998.