# Automatic Identification of Abnormalities in 12-lead ECGs Using Expert Features and Convolutional Neural Networks

Zhongdi.Liu
School of Electronic, Electrical and Communication Engineering University of Chinese Academy of Science
Beijing, China
zhongdiliu@icloud.com

Xiang'ao.Meng
School of Electronic, Electrical and Communication Engineering University of Chinese Academy of Science
Beijing, China
332909414@qq.com

Jiajia.Cui
School of Electronic, Electrical and Communication Engineering University of Chinese Academy of Science
Beijing, China
2997215325@qq.com

Zhipei.Huang*
School of Electronic, Electrical and Communication Engineering University of Chinese Academy of Science
Beijing, China
zhphuang@ucas.ac.cn

Jiankang.Wu
School of Electronic, Electrical and Communication Engineering University of Chinese Academy of Science
Beijing, China
jkwu@ucas.ac.cn

***Abstract*—Automatic identification of the rhythm/morphology abnormalities in ECGs has gained growing attention in various areas and remains a challenge. We propose an algorithm to classify 12-lead ECGs into 9 categories. We extracted expert features including generic features and specific features with statistics and physiology significance. Then a 17-layer Convolutional Neural Network (CNN) was proposed to detect deep features in ECGs. With these features we trained ensemble classifiers to predict labels. Experiment on the training set (5-fold cross-validation) reports 0.81 accuracy score.**

***Keywords-component; ECG classification; cardiac abnormalities; deep learning; Convolutional Neural Network***

## I. INTRODUCTION

Electrocardiograms (ECG) is a common non-invasive measurement that can reflect the physiology activities of heart. As one of the most important diagnostic tools, ECG signals have significant advantages for its convenience and efficiency. These years, with development of medical scale and smart wearable devices, patients can get their ECGs more ubiquitously. However, it's still a heavy task for doctors to handle such multifarious data. So robust automatic ECG classification algorithms are needed to relieve their pressure, meanwhile patients can evaluate cardiac condition themselves.

On account of the important role of ECG classification, a lot of related works have been done. It's natural to solve this problem with a view to expert features. Smisek extracted rhythm and morphology features from the separate beats, then support vector machine (SVM) and other machine learning models were used to do the prediction[1]. Goodfellow used template features, RPI features, and full waveform features as inputs of machine learning algorithm (XGBoost, tree-based, gradient boosting classifier)[2]. Deep learning has achieved great success in the field of computer vision, natural language processing, and speech recognition. In recent years, this powerful technique has gradually been used to solve problems in area of medical information and biomedical signal processing. A pioneering work was proposed by Rajpurkar[3], implemented a 34-layer CNN to detect arrhythmia and exceeded the average cardiologist performance in both recall and precision. Xiong developed RhythmNet, a 21-layer 1D Convolutional Recurrent Neural Network which was trained using 8,528 single lead ECG recordings from the 2017 PhysioNet/Computing in Cardiology (CinC) Challenge, to automatically classify ECGs of different rhythms including AF[4].

In this paper, we combine the two approaches above to achieve a more reliable classification system which can identify 8 typical abnormalities in ECGs. Section 2.1 describes the data and preprocessing method. The structure and key techniques of CNN are stated in section 2.2. In section 2.3 we explain significance and extraction of expert features. In section 2.4 the ensemble classifiers are described. Finally, the obtained results, discussions and conclusion are explained in Sections 3, 4 and 5, respectively.

## II. METHODS

### A. Data Description and Preprocessing

The challenge ECG recordings were collected from 11 hospitals. The training set contains 6,877 (female: 3178; male: 3699) 12 leads ECG recordings lasting from 6 s to just 60 s

978-1-5386-7413-0/18/$31.00 ©2018 IEEE
DOI 10.1109/SNSP.2018.00038

and the test set contains 2,954 ECG recordings with similar lengths. The dataset is provided by the China Physiological Signal Challenge 2018 (CPSC2018)[5].

In order to obtain a clearer original data, we preprocess the data by wavelet filtering. It mainly eliminates 50Hz power noise, the baseline shift and myoelectric noise caused by muscle tension in patients. In addition, there are some reasons for data invalidation (such as lead loss). We discard these invalid data and extract features on better performance signals.

### B. Convolutional Neural Network

CNNs are quite prevalent in the field of computer vision due to their advantages: translation invariance, parameter sharing and sparse connectivity. Recently some study proved that 1-D CNNs (1-Dimension Convolutional Neural Networks) are also effective tools for time sequence modeling[6]. One drawback of deep learning methods is that, the more complicated network structure are, the more difficult training networks task become. Rectified Linear Units (ReLU) proposed by Nair relief vanishing gradient problem[7]. Kaiming's Residual Networks (ResNet) use shortcut connections between layers to create identity mapping[8], making it possible to train deeper networks without accuracy falling. Nitish proposed Dropout[9], a powerful method to prevent overfitting, meanwhile Dropout could bring improvement of accuracy because of its effect of model ensemble. With constructive methods above, researchers can build Deep Neural Networks to model complicated problems.

In our work, a 17-layer 1-D ResNet (contains 17 layers of convolution followed by a fully connected layer) is designed to fit the training data. The network takes as input a time-series of raw ECG signal, and outputs label prediction. As length of the input signal is arbitrary, we copy signal and concatenate them to same length.

The network architecture is shown in Figure 1. We employ residual connections between every two convolutional layers. There are 8 residual blocks in networks, and every block includes 2 convolutional layers. All the filters in Convolutional layers have length of 15 and numbers of filters double per residual blocks. Meanwhile the inputs are downsampled by the factor of 2. When a residual block downsample the inputs, the residual connection also downsample the inputs by the same factor using a convolution operation with filters size of 1.

In each residual block we apply Batch Normalization[10], which has proved its effectiveness on accelerating network training. We use ReLU as activate functions. The final fully connected layer and softmax function output a probability distribution over 9 classes and the label corresponding to maximum probability is prediction of network.

Before training, method described in[11] is applied to initialize the weights of the Convolutional layers. We use Adam optimizer[12] to train the network with default parameters.

As our network above is an end-to-end model, after training, the final fully connected layer and softmax layer are removed, the outputs of 1-D maxpool layer are regarded as deep features. In the end 200 deep features are extracted.

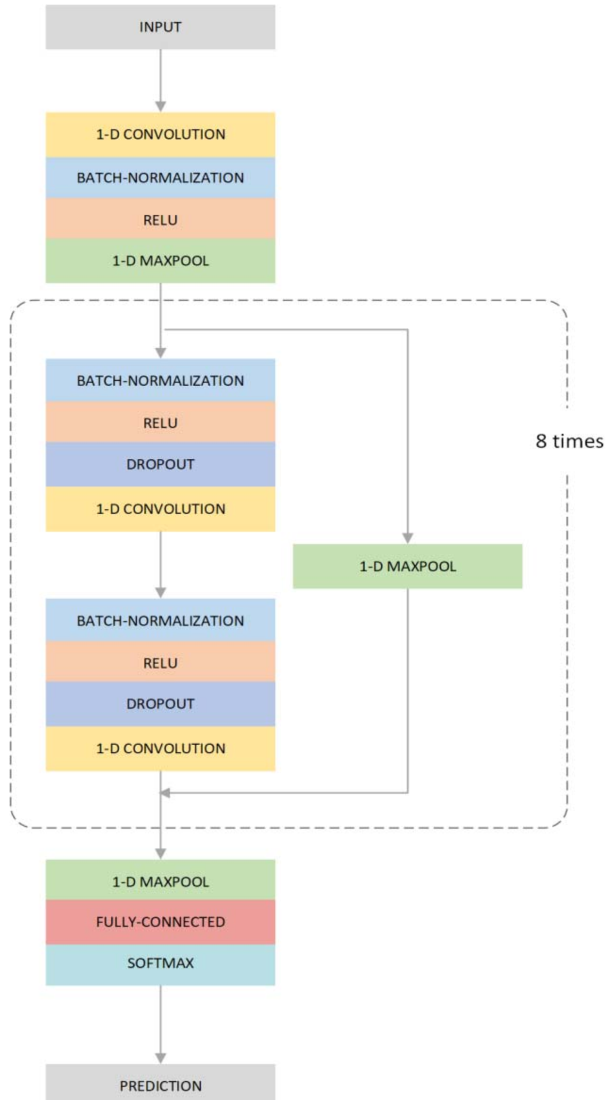


Figure 1. Network architecture, contains 11 layers of convolution followed by a fully connected layer and a softmax.

### C. Expert Features Extraction

In this section, we extract expert features including generic features and specific features with statistics and physiology significance.

#### 1) Generic features

Considering the repeability of features between each lead, we select the II lead for extraction of generic features in which diseases are obvious.[12].

We extract generic features in both time domain and frequency domain.

- Time domain features: The features in time domain can clearly reflect the characteristics of signals. We extract these features from statistics and morphology. Statistic features include count, mean, maximum, minimum, range, variance, skewness, kurtosis, percentile, mean HR, max HR, min HR, the variances of HR, etc. Amplitude of R, slope of R, duration of QRS, amplitude of P, interval of P-R, amplitude of T, interval of R-T are counted in morphological features, which is shown in Figure 2.

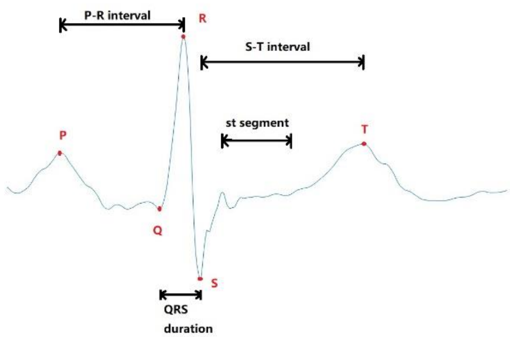


Figure 2. Morphological features in ECG signal, extracted in all 12 leads.

- Frequency domain features: We also transfer the signal to the frequency domain for feature extraction. These features are extracted from the statistical and energy aspects. We first implement FFT (Fast Fourier Transform) on ECGs, then we compute power, frequency band power and Shannon entropy[13].

2) *Specific features*

Considering the same anomaly performs differently in each leads, the doctor's advice is adopted to extract specific features. These features are not restricted to a certain lead, but are extracted in different leads for different anomalies. These features include amplitude of P wave， the number of peak between SR， S wave amplitude，R wave duration，st interval, duration of T, maximum of T， st segment features, etc. The details are described as follows.

- The amplitude of P wave: For atrial fibrillation (AF), the most obvious phenomenon is the disappearance of P waves and accompanied by irregular heart rate. Therefore, the amplitude of P wave in the V lead could be extracted. At the same time, the above heart rate variability also plays an important role in the identification of atrial fibrillation.
- The number of peaks between S-R, S wave amplitude and R wave duration: Since left bundle branch block (LBBB) and right bundle branch block (RBBB) would lead to the change of QRS waveform, the peak number between S-R is extracted (in aVL, Ⅰ, V1, V2 lead) to distinguish whether the R wave had the notch or not[14], and the S wave amplitude and R wave duration in V5V6 lead are also extracted.
- ST-T features: ST-T features include st interval, duration of T, and maximum of T and st segment features. On the one hand, RBBB and LBBB usually lead to morphological changes of the st-t segment, so it is necessary to extract the st-t characteristic information of V leads. On the other hand, the characteristic information of st segment is obviously needed in ST-segment change classification. So the st segment features such as representative value, maximum value, minimum value and mean value are extracted.

All the features are listed in Table I. We finally extracted 174 features from 12 leads, in which repeated features were extracted only once.

TABLE I. ALL FEATURES USED FOR CLASSIFICATION.

| Feature List | | |
|---|---|---|
| **Generic features** | **Time domain** | **Statistic features:** count, mean, maximum, minimum, range, variance, skewness, kurtosis, percentile, mean HR, max HR, min HR, the variances of HR<br>**Morphological features:** amplitude of R, slope of R, duration of QRS, amplitude of P, interval of P-R, amplitude of T, interval of R-T |
| | **Frequency domain** | **Statistic feature**: Shannon entropy<br>**Energy features:** power, frequency band power |
| **Specific features** | amplitude of P wave, number of peaks between S-R, S wave amplitude, R wave duration, st interval, duration of T, maximum of T, st segment features | |

## D. *Ensemble Classifiers*

Now we have expert features and deep features, which are concatenated to a feature vector, as is input of classifier.

We choose XGBoost[15] (eXtreme Gradient Boosting of decision trees) as classifier, which is an effective modified implement of Gradient Boosting. Study shows that ensemble classifiers are more accurate than individual classifiers in most instances[16], so we train several individual XGBoost classifiers, and ensemble them by average the predictions.

# III. RESULTS

The CSPC2018 requires to classify ECGs into 9 categories: (1) Normal, (2) Atrial fibrillation (AF), (3) First-degree atrioventricular block (I-AVB), (4) Left bundle branch block (LBBB), (5) Right bundle branch block (RBBB), (6) Premature atrial contraction (PAC), (7) Premature ventricular contraction (PVC), (8) ST-segment depression (STD), (9) ST-segment elevated (STE). The evaluation indicator for model performance on individual class is stated as equation (1), in which $i$ refers to index of class above. $N_{j,k}$ stands for counting of samples in $j$ class which are classified into $k$ class.

$$F_{1i} = \frac{2 \times N_{i,i}}{N_{i,x} + N_{x,i}} \tag{1}$$

The final challenge score is defined as equation (2):

$$F_1 = \frac{\sum_{i=1}^{9} F_{1i}}{9} \tag{2}$$

In order to choose CNN model of appropriate layers, several CNN models of different layers were implemented in our work. The performance is shown in Figure 3, obviously 17-layer CNN models performed best, which is selection in subsequent work.

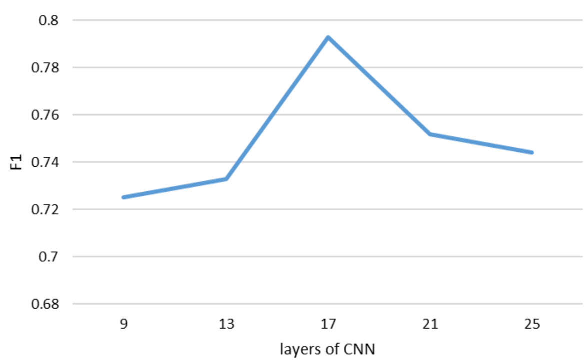


Figure 3. Performance of CNN of different layers.

Finally, our algorithm was evaluated by 5-fold validation. We mainly compared impact of using expert features and deep features. Results are shown in Table II.

TABLE II. RESULTS OF DIFFERENT FEATURES

| | $F_{11}$ | $F_{12}$ | $F_{13}$ | $F_{14}$ | $F_{15}$ | $F_{16}$ | $F_{17}$ | $F_{18}$ | $F_{19}$ | $F_1$ |
|---|---|---|---|---|---|---|---|---|---|---|
| **Expert** | 0.61 | 0.58 | 0.52 | 0.55 | 0.60 | 0.45 | 0.59 | 0.71 | 0.57 | 0.58 |
| **Deep** | 0.80 | 0.89 | 0.87 | 0.77 | 0.90 | 0.65 | 0.79 | 0.80 | 0.56 | 0.78 |
| **Expert+ Deep** | 0.82 | 0.91 | 0.87 | 0.87 | 0.91 | 0.63 | 0.82 | 0.81 | 0.60 | 0.81 |

It's apparent that when combining expert features and deep features, algorithm performs better. In addition using deep features alone could lead a result not much worse. However, only using expert features cannot bring a satisfying result, which means it's not enough to classify samples in high dimensional space with these features.

## IV. DISCUSSION

Due to its powerful fitting ability, deep learning methods have been used to solve problems in a wide range of area and they play the role of fast modeling. For ECG classification, our CNN model could provide a baseline performance. Further, combining expert features, our algorithm achieves capability closed to state-of-the-art level.

There are still some limitations in our work. On the one hand, as networks are driven by training data, performance on categories which contain less samples may be poor. It's noticed in Table II that $F_{16}$ and $F_{19}$ are apparently lower than others. With expert features, problems caused by sample imbalance can be alleviated but remain severe. On the other hand, considering that data filtering may cause a degree of loss in information, we choose to retain ECGs of all subjects, thus some bad data would reduce the quality of expert features extracted, affecting the classifying quality.

There is promotion space of our work. Other deep learning models could be implemented such as Recurrent Neural Networks, which may have better ability to model time series. The extraction algorithm of more morphological features of ECGs also needs to be further studied.

## V. CONCLUSION

In this paper, we implement a classifier using expert features and deep features for ECG classification. Experiments shows that using combination of features performs better than using them individually.

## REFERENCES


[1] R. Smisek, J. Hejc, M. Ronzhina, A. Nemcova, L. Marsanova, J. Kolarova, L. Smital, and M. Vitek, "Multi-stage SVM approach for cardiac arrhythmias detection in short single-lead ECG recorded by a wearable device," Physiological measurement, 2018-Aug-13, 2018.

[2] S. D. Goodfellow, A. Goodwin, R. Greer, P. C. Laussen, M. Mazwi, and D. Eytan, "Atrial fibrillation classification using step-by-step machine learning," Biomedical Physics & Engineering Express, vol. 4, no. 4, pp. 045005 (16 pp.)-045005 (16 pp.), July, 2018.

[3] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," ArXiv e-prints, 1707, http://adsabs.harvard.edu/abs/2017arXiv170701836R, [July 1, 2017, 2017].

[4] Z. Xiong, M. P. Nash, E. Cheng, V. V. Fedorov, M. K. Stiles, and J. Zhao, "ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network," Physiological measurement, 2018-Aug-13, 2018.

[5] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Yin Kwee, "An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection," Journal of Medical Imaging and Health Informatics, vol. 8, no. 7, pp. 1368-1373, //, 2018.

[6] M. D. Zeiler, and R. Fergus, "Visualizing and Understanding Convolutional Networks," ArXiv e-prints, 1311, http://adsabs.harvard.edu/abs/2013arXiv1311.2901Z, [November 1, 2013, 2013].

[7] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 2010, pp. 807-814.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," ArXiv e-prints, 1512, http://adsabs.harvard.edu/abs/2015arXiv151203385H, [December 1, 2015, 2015].

[9] N. Srivastava, G. Hinton, A. Krizhevsky, Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research, 15, http://adsabs.harvard.edu/abs/2015arXiv151203385H, 2014].

[10] S. Ioffe, and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," ArXiv e-

prints, 1502, http://adsabs.harvard.edu/abs/2015arXiv150203167I, [February 1, 2015, 2015].

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," ArXiv e-prints, 1502, http://adsabs.harvard.edu/abs/2015arXiv150201852H, [February 1, 2015, 2015].

[12] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," ArXiv e-prints, 1412, http://adsabs.harvard.edu/abs/2014arXiv1412.6980K, [December 1, 2014, 2014].

[13] H. Shenda, W. Meng, Z. Yuxi, W. Qingyun, S. Junyuan, L. Hongyan, and X. Junqing, "ENCASE: An ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks," 2017 Computing in Cardiology (CinC), pp. 4 pp.-4 pp., 2017, 2017.

[14] D. G. Strauss, and R. H. Selvester, "The QRS complex—a biomarker that "images" the heart: QRS scores to quantify myocardial scar in the presence of normal and abnormal ventricular conduction," Journal of Electrocardiology, vol. 42, no. 1, pp. 85-96, 2009/01/01/, 2009.

[15] T. Chen, and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," ArXiv e-prints, 1603, http://adsabs.harvard.edu/abs/2016arXiv160302754C, [March 1, 2016, 2016].

[16] T. G. Dietterich, "Ensemble Methods in Machine Learning," Multiple Classifier Systems. pp. 1-15.