Large-scale Classification of 12-lead ECG with Deep Learning

Yu-Jhen Chen*, Chien-Liang Liu*, Vincent S. Tseng†, Yu-Feng Hu‡, Shih-Ann Chen‡
*Department of Industrial Engineering and Management
National Chiao Tung University
lycka.iem04g@nctu.edu.tw, clliu@mail.nctu.edu.tw
†Department of Computer Science
National Chiao Tung University
vtseng@cs.nctu.edu.tw
‡Heart Rhythm Center, Division of Cardiology, Department of Medicine
Taipei Veterans General Hospital
huhuhu0609@gmail.com, epsachen@ms41.hinet.net

***Abstract*—The 12-lead Electrocardiography(ECG) is the gold standard in diagnosing cardiovascular diseases, but most previous studies focused on 1-lead or 2-lead ECG. This work uses a large data set, comprising 7,704 12-lead ECG samples, as the data source, and the goal is to develop a classification model for six common types of urgent arrhythmias. We consider the characteristics of multivariate time-series data to design a novel deep learning model, combining convolutional neural network (CNN) and long short-term memory (LSTM) to learn feature representations as well as the temporal relationship between the latent features. The experimental results indicate that the proposed model achieves promising results and outperforms the other alternatives. We also provide brief analysis about the proposed model.**

***Index Terms*—12-lead ECG, Classification Model, Deep Learning, CNN, LSTM**

I. Introduction

The Electrocardiogram (ECG) provides objective information about the electrical activity of the heart over a period of time using electrodes placed over the skin. Urgent arrhythmias such as atrial fibrillation (AF), ventricular tachycardia and atrioventricular (AV) block could be diagnosed by ECG. However, the substantial mis-diagnoses of these arrhythmias were still common, even by a trained cardiologist. The overloaded work of the physician further worsens the overlook of emergent arrhythmias. Automatic classification of cardiac arrhythmias is a valuable tool to screen the arrhythmias and assist in the physician to make the precise diagnosis and reduce working time.

Although many researchers have devoted to developing classification models for ECGs, most of them used 1-lead or 2-lead ECG as the data source [6], [12], [15] as it is difficult to obtain large-scale 12-lead ECG data from medical institutions. However, the mostly utilized test to detect cardiovascular diseases in the clinical settings is the standard 12-lead ECG, which is made up of the three standard limb leads, the augmented limb leads and the six precordial leads. The 12-lead ECG is more complex and comprehensive than the 1-lead or 2-lead ECGs, making it difficult to apply present models with few leads to the 12-lead ECG.

The last decade has witnessed the great success of deep learning as it has brought revolutionary advances in many application domains, including computer vision, natural language processing and signal processing. The key idea behind deep learning is to consider feature learning and classification in the same network architecture, and use backpropagation to update model parameters to learn discriminative feature representations. More importantly, many novel deep learning methods have been devised and improved classification performance significantly [5], [11], [14]. For example, ResNet [5] applied the concept of residual blocks, which could efficiently deal with gradient vanishing and exploding problems that often exist in deep neural networks. Google inception model [13], [14] applies various convolutional kernel sizes, such as 1×1, 3×3 and 5×5, to obtain different feature maps.

This work proposes to use deep learning to develop a classification of cardiac arrhythmias with the 12-lead ECG. We used the real-world data collected from a clinical institution as the data source, which comprised 7,704 samples. The aim is to classify six common types of urgent arrhythmias, including AF, atrial flutter, complete AV block, junctional rhythm, sinus node disease and WolffParkinson-White syndrome (ventricular pre-excitation). The 12-lead ECG is the gold standard in medical examination for cardiac diseases with the universal format through all clinical institutions. As a result, the proposed method could be easily generalized and disseminated to other 12-lead ECG data sources.

II. Related Work

The ECG comprises three main components, including the P wave, QRS complex and T wave. The P wave represents the depolarization of the atria, while the QRS complex and the T wave denote the depolarization and repolarization of the ventricles, respectively. To apply machine learning algorithms

This study was supported in part by the Ministry of Science and Technology, Taiwan, under Grant no. MOST 107-2221-E-009-109-MY2 and MOST 107-2218-E-009-005.

978-1-7281-0848-3/19/$31.00 ©2019 IEEE

to the classification of cardiac arrhythmias, the ECG signals should be represented as feature vectors. It is apparent that these feature vectors are crucial to the classification performances of the machine learning algorithms, and most of the reported approaches rely on extracting hand-crafted features from the signals.

Hand-crafted features are normally obtained from domain knowledge or signal processing techniques. The aforementioned components of the ECG signal could be used to identify ECG features, such as peaks, intervals and segments. Therefore, several research studies focused on the detection of these features, so that the classification algorithms could benefit from these ECG features. Besides, the ECG is a time-series sequence, so signal processing techniques such as discrete wavelet transformation, discrete Fourier transformation and continuous wavelet transformation are also commonly used to extract meaningful features. The survey conducted by Jambukia et al. [6] summarized the features extracted from ECG and signal processing techniques. Once the feature extraction process is completed, different machine learning algorithms such as support vector machines (SVM) [9], [15], neural networks [12] and logistic regression [3] could be applied to learn a predictive model from available training data.

Extracting hand-crafted features is a time-consuming and expensive task, and a small variations in the values of extracted features may have huge impact on the classification performance over large data sets. In recent years, many researchers have proposed to use deep learning to directly learn an end-to-end classification model without hand-crafted features. The goal of deep learning is to jointly learn data representations and model parameters from a collection of data. The realization of deep learning relies on deep neural networks, which involve a cascade of many layers of nonlinear processing units for feature extraction and transformation. In the architecture, the output of a specific layer becomes the input of the next layer, and the layer architecture makes it possible to learn a hierarchical representation.

Among these deep learning architectures, the CNN has achieved a breakthrough improvement on image classification. Xiong et al. [17] designed a 16-layer CNN model training on single-lead-ECG dataset to classify three ECG symptoms and other types: normal rhythm, AF rhythm, other rhythm and noisy recordings. Limam and Precioso [10] used two independent CNNs to extract relevant patterns, in which one was from the ECG and the other was from the heart rate. The two extracted feature maps were merged into a recurrent neural network (RNN) accounting for the sequence of the extracted patterns. Hannun et al. [4] developed a 34-layer CNN to diagnose abnormal heart rhythms. The ECG involves complex patterns, explaining why a deep CNN was used to learn feature representations. Similarly, Karasawa et al. [7] focused on the diagnosis of Alzheimers disease, and proposed a deep 3D CNN with 39 layers, making it possible to learn good feature representations and obtain promising results.

III. Proposed Method

The input data in this work is 12-lead ECG sequences, belonging to multivariate time-series data. It is expected that the correlation between the leads and the sequence ordering in the time series data are crucial to model performance. To extract discriminative features while retaining the temporal information encoded in the time-series data, this work proposes a novel architecture combining CNN and LSTM to classify the input 12-lead data. Figure 1 shows the proposed architecture. The proposed model could not only learn 12-lead information individually, but also benefit form 12-lead coordinated information, giving a base to imitate what medical experts do when making diagnosis with ECG.

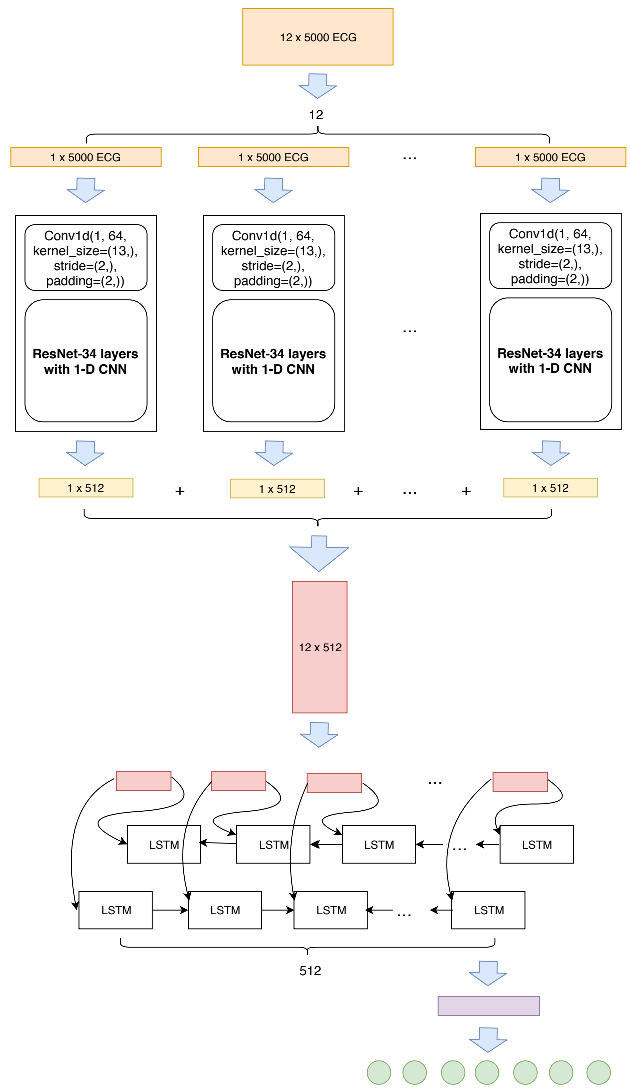

Fig. 1. Proposed Deep Learning Architecture

The input 12-ECG comprises 12 leads, each of which comprises 5,000 data points. In our proposed model, we treat each lead as the input of the CNN model, so the input size for the CNN model is 1×5000. Then, we use ResNet with 34 layers to extract features for each lead. As compared with original ResNet, this work uses 1D convolution rather than 2D convolution as our goal is to learn feature representations for each lead, which is a 1D vector. Therefore, we apply 1D convolution and pooling over the input data in our ResNet model. Additionally, we eliminate the fully connect layer as used in ResNet, since the purpose of this step is to extract features instead of making prediction.

Once the feature learning process is completed, we could obtain a feature vector of size 1×512 for each lead. The 12 feature vectors are concatenated to form a matrix of size 12×512, which is the input of LSTM. This step is to use LSTM to learn the temporal relationship between the extracted features. In the implementation, we use a bi-directional LSTM as presented in Figure 1. The output of the LSTM is connected with a fully connected layer to make prediction. The goal is to perform classification on 12-lead ECG data, so we use cross entropy loss as the loss function.

IV. Experiments

A. Dataset

The dataset used in this work comprises 7,704 12-lead ECG samples, which were collected from Taipei Veterans General Hospital Division of Cardiology. De-identification is applied to the data to protect patient privacy. This work focuses on emergent heart diseases, and considers six urgent types of diagnoses in the experiments, including Atrial fibrillation, Atrial flutter, Complete AV block, Junctional rhythm, sinus node disease, and WPW. Besides these six diseases, the data that do not belong to any of the six classes are grouped into a class called “other”. We focus on multi-class problem, indicating that each 12-lead ECG sample belongs to one class. Summary of the dataset is presented in Table I.

TABLE I
Summary of the dataset

Class	Number of samples
Atrial fibrillation	2,155
Atrial flutter	1,270
Complete AV block	328
Junctional rhythm	280
sinus node disease	987
WPW	684
other	2,000
total	7,704

B. Experimental Setting

The proposed model comprises CNN and LSTM. In the implementation, we use 1D convolution in the ResNet-34 to extract discriminative features. As for LSTM model, we use bidirectional LSTM with one layer, and the purpose is to pass feature information back and forth to capture temporal relationship. The input size for the bidirectional LSTM is 12, since we would like each LSTM cell to view the feature maps learned from each lead of ECG. Additionally, the number of time steps for the bidirectional LSTM is 512. The batch size is 16, and learning rate is 0.0001. Finally, We use Adam [8] as our optimizer.

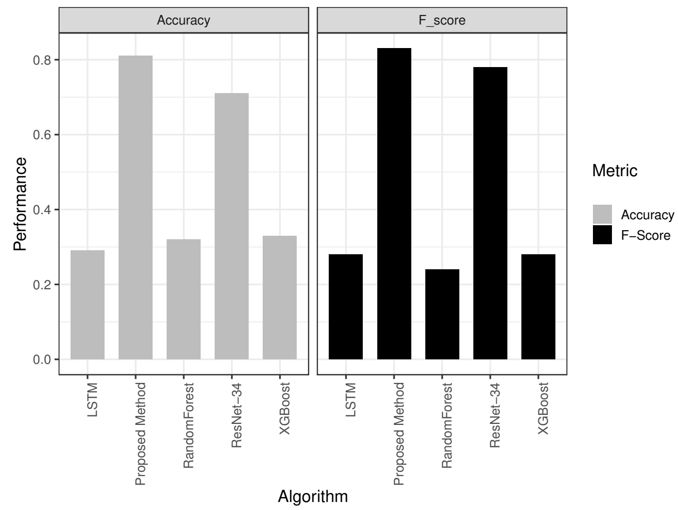

Fig. 2. Experimental Results

C. Experimental results

We use 5-fold cross validation to conduct experiments, and compare the proposed model with several alternatives, such as XGBoost [2] and Random Forest [1]. Additionally, the comparison methods involve the model with only LSTM or 1D-CNN RestNet-34, and these two models are the baseline models. The experimental results are presented in Figure 2, which involves two evaluation metrics, accuracy and macro F1-score, in which the final F1-score is the average of all classes. The parameter setting for LSTM method is the same as our LSTM model. As for the other comparison methods, we try to use cross validation to obtain the best settings.

As shown in Figure 2, the proposed model outperforms the other alternatives. Although both XGBoost and random forest are state-of-the-art methods, the experimental results indicate that they fail to perform well on 12-lead ECG data.

This work focuses on end-to-end models, so the input data are 12-lead ECG without any feature extraction process. As a result, each input sample is a matrix of size 60,000. Even though the number of data samples is large, the number of data samples is still much less than the dimension size. In this case, the machine learning algorithms are easy to suffer from the problem of curse of dimensionality. We conclude that is the main reason why these two algorithms fail to perform well. We have tried to apply principal component analysis (PCA) to reduce the dimensionality of the input data, but the performances are still poor as PCA is a linear algorithm for dimensionality reduction, whereas the 12-ECG signals are expected to have nonlinear structure. The experimental results indicate that using traditional machine learning algorithms requires to extract features before model training.

In contrast, the performance of 1D-CNN RestNet-34 works well as it could learn discriminative feature representation from data, giving a base to classify 12-ECG data well. The proposed model combines 1D-CNN RestNet-34 and LSTM together, making it be able to benefit from the features learned from CNN, as well as the temporal relation captured by LSTM.

As for LSTM, the input data is a matrix of size 12×5000, explaining why we set the number of time steps for LSTM to be 5,000. The LSTM may run into the problem of vanishing gradients and exploding gradients when dealing with a very long sequence. Thus, only using LSTM without CNN fails to perform well on the prediction of 12-lead ECG.

D. Discussion

In 5-fold accuracy cross validation, the average accuracy for our proposed model is 0.81, and the performance differences among the folds are minor. The confusion matrix of the best fold is presented in Table II.

TABLE II
CONFUSION MATRIX

	Atrial fibrillation	Atrial flutter	Complete AV block	Junctional rhythm	sinus node disease	WPW	other
Atrial fibrillation	**376**	32	7	10	8	2	15
Atrial flutter	17	**212**	3	1	3	2	9
Complete AV block	1	4	**47**	5	4	0	1
Junctional rhythm	3	0	5	**40**	8	1	3
sinus node disease	5	3	6	6	**156**	2	25
WPW	2	1	0	0	1	**136**	14
other	4	7	0	6	37	5	**305**
total	408	259	68	68	217	148	372

As shown in Table II, the average accuracy for our proposed model is 0.81 (Sensitivity: 0.82, specificity: 0.97), indicating that our proposed model could provide accurate prediction on 12-lead ECG data samples. Among the six classes in the experiments, three of them still have the room of improvement. The main reasons could be attributed to the limited numbers of data samples for these three classes or the imbalanced class distribution in the data. The multiple diagnoses of arrhythmias in the same patient were possible, meaning that those patients may suffer from more than one arrhythmia. The present model could provide a single prediction outcome, so that is another possible reason.

The diagnostic sensitivity and specificity of the present model were higher than the accuracy of the interpretation of ECGs by general practitioners and cardiologists [4], [16]. The mean negative prediction rate was 0.97 (lowest to highest: 0.94 to 0.99). This finding suggested the proposed model is potentially applied as the powerful screening and validating system to exclude these urgent arrhythmias. The 12-lead ECG is the most commonly used ECG tool. Its wide coverage in the medical institutes from nursing centers to medical centers implied the translation of the present model would have substantial clinical impact than other ECG recording system (e.g., 1 lead ECG patch or holters). Awaiting the clinical trials and future incorporation into the clinical settings, the proposed system might enhance the prompt diagnosis and management of these urgent arrhythmias. This is especially helpful in the non-cardiology specialists, general practitioner, resident trainees, or nurse as the delayed and misdiagnosis are common due to lacking of domain knowledge.

V. CONCLUSION

This work proposes to use deep learning technique to deal with the classification of 12-lead ECG. In our proposed model, we extend RestNet to devise a CNN architecture to learn good feature representations. Then, we use a bidirectional LSTM to retain the temporal relationship among the latent variables learned from CNN. The experimental results indicate that our proposed model shows promising results on a large data set. The future work is to consider imbalanced characteristic in the model to improve the classification performances for some rare diseases.

REFERENCES

[1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM, 2016.

[3] M. A. Escalona-Morán, M. C. Soriano, I. Fischer, and C. R. Mirasso. Electrocardiogram classification using reservoir computing with logistic regression. *IEEE Journal of Biomedical and health Informatics*, 19(3):892–898, 2015.

[4] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65, 2019.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[6] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati. Classification of ecg signals using machine learning techniques: A survey. In *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in*, pages 714–721. IEEE, 2015.

[7] H. Karasawa, C.-L. Liu, and H. Ohwada. Deep 3d convolutional neural network architectures for alzheimers disease diagnosis. In *Asian Conference on Intelligent Information and Database Systems*, pages 287–296. Springer, 2018.

[8] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] Q. Li, C. Rajagopalan, and G. D. Clifford. A machine learning approach to multi-level ecg signal quality classification. *Computer methods and programs in biomedicine*, 117(3):435–447, 2014.

[10] M. Limam and F. Precioso. Atrial fibrillation detection and ecg classification based on convolutional recurrent neural network. *Computing*, 44:1, 2017.

[11] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu. Time series classification with multivariate convolutional neural network. *IEEE Transactions on Industrial Electronics*, 66(6):4788–4797, 2019.

[12] M. Mitra and R. Samanta. Cardiac arrhythmia classification using neural networks with selected features. *Procedia Technology*, 10:76–84, 2013.

[13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[15] E. D. Übeyli. Ecg beats classification using multiclass support vector machines with error correcting output codes. *Digital Signal Processing*, 17(3):675–684, 2007.

[16] H. C. van Weert. Diagnosing atrial fibrillation in general practice. *BMJ: British Medical Journal*, 335(7616):355, 2007.

[17] Z. Xiong, M. K. Stiles, and J. Zhao. Robust ecg signal classification for detection of atrial fibrillation using a novel neural network. *Computing*, 44:1, 2017.