Received June 17, 2020, accepted June 26, 2020, date of publication July 2, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006707

Interpretation of Electrocardiogram (ECG) Rhythm by Combined CNN and BiLSTM

XUE XU[1,2], SOHYUN JEONG[2], AND JIANQIANG LI[1], (Member, IEEE)
[1]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[2]Hebrew SeniorLife and Harvard Medical School, Marcus Institute for Aging Research, Boston, MA 02131, USA

Corresponding author: Jianqiang Li (lijq@szu.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant U1713212 and Grant 61572330, in part by the Natural Science Foundation of Guangdong Province-Outstanding Youth Program under Grant 2019B151502018, in part by the Technology Research Project of Shenzhen City under Grant JSGG20180507182904693, and in part by the Public Technology Platform of Shenzhen City under Grants GGFW2018021118145859.

ABSTRACT Computer-aided detection and diagnosis in ECG signals for heart diseases are gaining increasing attention. However, developing and selecting the highly performing diagnostic model suitable for clinical implications is still challenging. In this paper, we proposed a combined network of convolutional neural network (CNN) and Recurrent Neural Network (RNN), designed for the classification of ECG heart signals for diagnostic purpose. The proposed network consists of 2 convolutional layers with 5×5 kernels and ReLU activations, followed by 4 residual blocks, 2 bidirectional long short-term memory (biLSTM) layers, as well as 2 fully connected layers. Each residual block involved the structure of a Squeeze-and-Excitation Network (SENet) with lightweight property to recalibrate the feature map of the network. The last dense layer has 5 outputs, equivalent to the classes considered: non-ectopic beats, supraventricular ectopic beats, ventricular ectopic beats, fusion beats, and unknown beats. To train and evaluate the combined CNN and RNN, we transferred the knowledge acquired on beat classification tasks in 2017 PhysioNet/CinC Challenge to that in PhysionNet's MIT-BIH dataset. The developed network achieved a recognition sensitivity of 95.90%, accuracy = 95.90% and specificity = 96.34% with classification time of single sample = 6.23 s in detecting 5 ECG classes. A comparative analysis proved the high performance of the proposed combined CNN and RNN against previous methods, demonstrating the potential of our proposed network in the analysis of beat patterns. The proposed model can be applied in cloud computing or implemented in mobile devices to evaluate cardiac health with the highest precision.

INDEX TERMS Electrocardiogram (ECG), classification model, convolutional neural network (CNN), recurrent neural network (RNN).

I. INTRODUCTION

Electrocardiogram (ECG) is the most widely used diagnostic tool in cardiology [1]. ECG uses external electrodes to measure the electrical conduction signals of the heart and record them as characteristic lines. These lines allow the axis, rate, and rhythm, as well as the amplitudes of specific parts of the heart (e.g., the P wave, PR interval, QRS complex, ST segment) to be examined. Understanding and interpreting the ECG recordings are crucial to increase the accuracy of diagnosis and introduce timely management to patients with abnormal heart rhythms, heart attack, and enlarged hearts [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou.

Despite the evolution of consensus statements and guidelines regarding the ECG interpretation enhanced identification of cardiovascular pathology [1], [2], [4], manual interpretation of ECG is still time-consuming and cumbersome. The categorization of different waveforms and morphologies into the standardized signals has a few challenges due to diversity amongst the ECG features, the individuality of the ECG patterns, and variability in ECG waveforms of patients [5]. Thus, developing the most appropriate classifier that is capable of classifying arrhythmia in real-time becomes a critical issue in ECG arrhythmia classification. ECG classification includes preprocessing, feature extraction, feature normalization, and classification.

In the recent decade, multiple machine learning algorithms and approaches have been developed to interpret accurate results from ECG and diagnose various cardiac conditions, such as fuzzy-based machine learning [6], rough set theory [7], support vector classifiers [8], and neural networks [9]–[11]. Based on the neural networks, deep learning with multiple hidden layers has been further developed to automatically extract the hidden signatures from the raw data and use this knowledge for the classification, instead of manually extracting the features from the raw ECG data with hand-crafted techniques in the machine learning techniques.

In neural network models, besides two representative training methods (back-propagation (BP) algorithm and deep belief network (DBN)) in deep learning-based ECG diagnosis [12], convolutional neural networks (CNN) and recurrent neural networks (RNN) have been widely used in ECG classification. A 5-layer CNN network was constructed based on the small size of a public PAF prediction challenge database to identify individuals with paroxysmal atrial fibrillation (PAF) [13]. By using the MIT-BIH arrhythmia database, 16-layer 1D-CNN was designed as well to classify 17 ECG classes of cardiac arrhythmias (normal sinus rhythm, 15 cardiac arrhythmias, and pacemaker rhythm). However, despite of the high accuracy (91.33%) and low computational complexity (classification time per single sample of 0.015 sec), the experiment was conducted only based on 3600 10-s ECG samples, with the most ECG signal fragments of 283 for the normal sinus rhythm class [14]. Similarly, a 3-layer deep genetic ensemble of classifiers was constructed to detect the 17 arrhythmia ECG classes, with improved accuracy of 99.37% [15]. However, the complex classifier needed feature extraction and longer training and optimization time. In addition, a 10-layer deep CNN model was designed to identify patients with myocardial infarction (MI) by using ECG records of 651 normal and 651 MI samples in Physiobank ECG database, with accuracy of over 99% [16].

However, it is worth noting that ECG is a time series signal that reflects the electrical activity of the heart. The signal consists of a series of repetitive and stereotyped complex waveforms with an obvious frequency of approximately 1 Hz. Although CNN has high computational efficiency, it is more suitable for spatial data such as images but RNN is more sensitive to sequential time series data. Therefore, a 3-layer RNN model was constructed to detect Power-Line Interference (PLI; one of the main noise components in ECG) resided in ECG signals [17]. Furthermore, a Long Short-Term Memory (LSTM) RNN was constructed to classify the normal and abnormal beats in an ECG, with accuracy of 88.1% [18]. However, RNN is difficult to train because it requires memory-bandwidth-bound computation.

Considering the advantages and disadvantages of CNN and RNN, we here proposed a hypothesis that merging CNN and RNN into a single network could improve prediction and accuracy of ECG classification and be more suitable for industry and clinic application [19].

Moreover, it is noteworthy that Zhu et.al merged a bidirectional LSTM and CNN to generate synthetic ECG data that agreed with existing clinical data so that the features of patients with heart disease could be retained [20]. Usually, not enough medical data is available to construct a deep learning neural network model, so the development of the concept of knowledge transfer [21], i.e., utilizing knowledge acquired for one task to solve related ones has been developed. Salem *et al.* [22] used the knowledge learned from images in the ImageNet dataset, consisting of many classes of images such as animals and objects, to the ECG domain by a deep neural network (DenseNet). Similarly, Xiao *et al.* [23] used a pretrained Google Inception V3 model based on the ImageNet dataset to retrain a CNN model, and thus detected ischemic ST changes in ECG. Also, Wu *et al.* [24] employed AlexNet weight initialization trained on the ImageNet and trained 2D-CNN to classify ECG signals into normal and abnormal beats. The above systems show that transfer learning strategies are applied on different pre-trained neural networks to carry out final classification. In this work, we aimed to construct a pre-trained model for identification of Atrial Fibrillation (AF) with the 2017 PhysioNet/CinC Challenge dataset for transfer learning. Then, we were to use the weights saved from the pre-trained model as the initial weights to classify 5 different beat categories in accordance with Association for the Advancement of Medical Instrumentation (AAMI) EC57 standard by combining CNN and RNN (Figure 1). We expected our model could support clinical diagnosis and application through high performance with small model size.

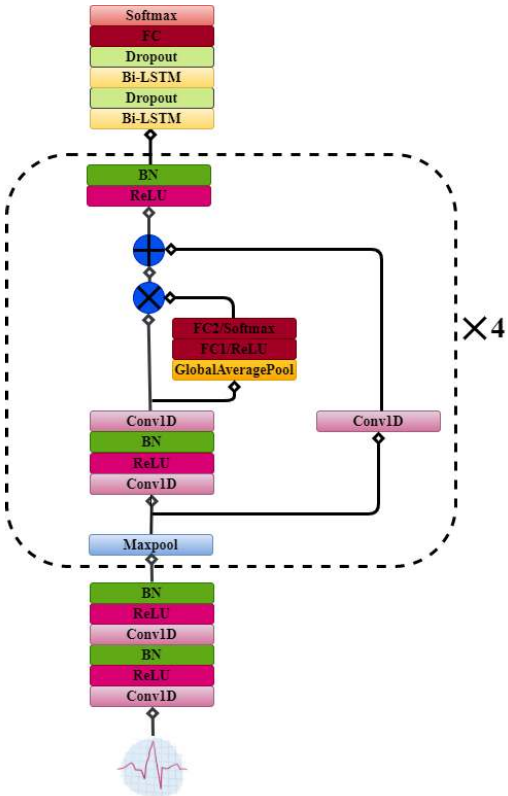

FIGURE 1. Architecture of our proposed model.

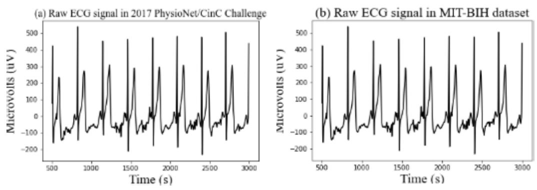

FIGURE 2. Real ECG signal. (a) and (b) show raw signals in 2017 PhysioNet/CinC Challenge dataset and MIT-BIH Dataset, respectively.

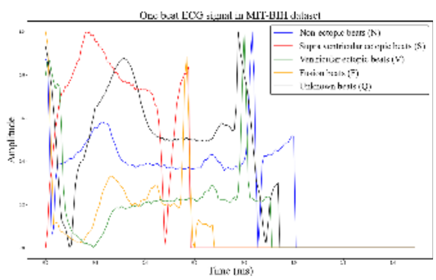

FIGURE 3. Denoised ECG signal. The ECG signal shows characteristics of 5 different beats in MIT-BIH Dataset.

II. METHODS

In this section, we first described details of the datasets involved in the study and data preprocessing steps, and then introduced our proposed deep learning method. Before the description of proposed classification network, we also presented the characteristics of the input data to better understand our proposed model.

A. DATASETS

The training set of 2017 PhysioNet/CinC Challenge [25] consists of 8528 one lead ECG recordings lasting from 9s to 61s at the sampling rate of 300Hz, involving normal rhythm, AF rhythm, other rhythm, and noisy recording. The dataset was used to train initial weights for transfer learning. The MIT-BIH Arrhythmia Dataset [26] was used for constructing the neural network in our work. The MIT-BIH dataset contains 48 ECG recordings from 47 subjects recorded at the sampling rate of 360Hz. Table 1 shows the mapping of 5 different beat categories for this dataset following AAMI EC57 standard. Figure 2 a and b show raw ECG signals captured with a sampling frequency of 300 Hz and 360 Hz in 2017 PhysioNet/CinC Challenge and MIT-BIH Arrhythmia Dataset. After applying denoising and beat detection techniques, Figure 3 shows the ECG signal with 5 categories in MIT-BIH Arrhythmia Dataset.

B. DATA PREPROCESSING

Before training, preprocessing of ECG signal was performed, in order to remove the baseline wander, motion artifacts, and other interruptions of the original recorded signal [27]. Our preprocessing strategy was simple, without any form of filtering or any processing that made any assumption about the signal morphology or spectrum. Five steps were involved in the preprocessing strategy for the ECG data.

TABLE 1. Mapping of 5 AAMI EC57 categories 25 with heartbeats in MIT-BIH Arrhythmia Dataset.

AAMI EC57:1998	Annotations of MIT-BIH heartbeats
Non-ectopic beats (N)	·Normal beats ·Left bundle branch block beats ·Right bundle branch block beats ·Nodal (junctional) escape beats ·Atrial escape beats
Supra ventricular ectopic beats (S)	·Aberrated atrial premature beats ·Supraventricular premature beats ·Atrial premature contraction ·Nodal (junctional) premature beats
Ventricular ectopic beats (V)	·Ventricular flutter wave ·Ventricular escape beats ·Premature ventricular contraction
Fusion beats (F)	·Fusion of ventricular and normal beats
Unknown beats (Q)	·Paced beats ·Unclassifiable beats ·Fusion of paced and normal beats

(1) Down-sampling: The ECG signals in 2017 PhysioNet/CinC Challenge and MIT-BIH Arrhythmia dataset was down-sampled from 300 Hz and 360Hz to 125Hz, respectively.
(2) Normalization: The ECG signals were normalized to have values between 0 and 1.
(3) R-peak and T episode detection: The ECG R-peaks were detected based on a threshold of 0.9 on the normalized value of the local maximums. T episodes

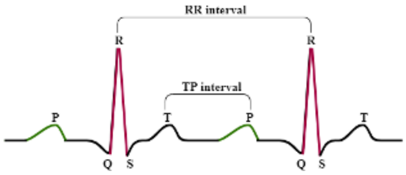

FIGURE 4. ECG description of PQRST waveforms.

(i.e., intervals during which the ECG exhibits significant T-wave changes) were further identified by the median of R-R time intervals.

(4) Beat extraction: ECG beat signals were extracted from the ECG T episodes. Reference annotation files were used to label each beat.

(5) Zero padding: Since the input length of beats for our model has to be identical, we set the maximal length to 10s for the 2017 PhysioNet/CinC Challenge dataset and 30s for the MIT-BIH dataset, respectively. Longer beats are cut at 10s/30s and shorter ones are padded with 0.

C. ARCHITECTURE OF PROPOSED MODEL

In order to design the optimal architecture and configuration of a deep learning model suitable for solving ECG heartbeat classification tasks, it's necessary to fully understand the characteristics of ECG signals. As shown in Figure 4, in a beat cycle, a normal ECG signal is mainly composed of P wave, QRS wave group, and T wave. These different waveforms contain very rich information. For example, the PR interval reflects the atrioventricular conduction time, which is also called the time limit of the cardiac cycle; the ratio of the PR interval to the RR interval reflects paroxysmal supraventricular tachycardia. Recognition of the waveforms in the ECG signal and extraction of the corresponding features are important steps to conduct the ECG heartbeat classification task. In fact, before application of artificial intelligence algorithms in ECG classification tasks, wavelet transform method [28], [29] was generally used to identify QRS complex in the ECG signal, and to determine the characteristics and positions of other typical complexes based on the QRS complex. Also, it helped to perform the feature extraction operation, to obtain the time domain or frequency domain features of detected complexes. The ECG features obtained from the wavelet transform were finally the inputs to traditional machine learning classification algorithms, i.e., support vector machines (SVM) and random forests, which probably resulted in weak generalization and discrimination performance. To avoid the disadvantages of manually selected features, convolution kernels of CNNs were developed [30]. At present, CNNs have been successfully applied in many ECG signal analysis and processing tasks [13], [14], which encourages the application of CNNs and their based networks to extract and classify ECG signal features. However, 1D CNN obtains global information mainly by aggregating local information which only involves the current input. In analyzing and processing time-series ECG data, a single CNN structure confronts the problem in modeling long sequence information. Therefore, to overcome this performance limitation in using 1D CNN, it is necessary to formulate an appropriate algorithm structure framework and training strategy so the model can reduce the running time and improve the accuracy rate.

1) ARCHITECTURE OVER VIEW

Based on the previous analysis, we designed a deep learning algorithm framework shown in Figure 1, which was used to implement the learning task from sequence to sequence. The proposed algorithm consists of two parts: a CNN and an RNN. First, the raw time series ECG signal was used as input into the CNN to extract sequence features, and then these features were employed as input into the RNN to perform heartbeat classification tasks. In addition, in the training process of the combined model, we introduced transfer learning, which used the weight parameters pre-trained on the 2017 PhysioNet/CinC Challenge dataset to initialize our network, thereby helping to find the global optimal solution, and finally improving the classification performance of our model. In the following section, we described CNN, RNN and transfer learning of the proposed network in detail.

2) DESCRIPTION OF CNN LAYERS

The raw ECG signal was passed through 2 consecutive 1D convolutional layers and was then input to the convolution module in the dashed box shown in Figure 1. For the convolution module, there were 4 modules in total, and each module had 3 convolutional layers. The filter size of the convolutional layer in the model was all 5. The number of convolution kernels of the first and the second convolutional layers was 32, while the numbers of convolution kernels for the four modules in the dashed box were 64, 96, 128, and 160, respectively. For the module in the dashed box, we adopted a residual connection method [31] similar to the deep residual network to make the model fully use the global context information, and introduced a Squeeze-and-Excitation (SE) module [32] to improve the sensitivity of our model for channel features. When the convolutional network performed convolution layer by layer, the obtained features were transferred from details (textures, lines, etc.) to abstractions (parts of the entire object, rough outlines, etc.). The residual block could catch different information from the feature sets extracted by different convolutional layers, thus effectively enhanced the discrimination capability of features and avoided a large increase in model parameters. Furthermore, the residual block not only provided feature reuse but also enhanced gradient flow. The activation function of all the CNN layers was ReLU function. Also, we added a batch normalization layer after each convolutional layer to avoid overfitting during training.

3) DESCRIPTION OF RNN LAYERS

As shown in Figure 1, after the feature extraction by CNN layers with raw ECG signal as input, the features were

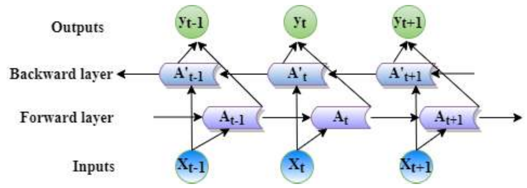

FIGURE 5. Architecture of bidirectional LSTM.

input to the RNN layer for classification. For the ordinary CNN model, the signal of neurons in each layer can only be propagated from the lower layer to the upper layer, and its processing for samples is independent at every moment. Moreover, CNN focuses more on the spatial correlation of features, which makes it relatively difficult to construct a model using time series data. While for the RNN model, the output of the neurons in each layer can directly affect itself in the next time period, so RNN can be used to describe the output of continuous state over time, and is more suitable for the analysis tasks on time series data such as speech, text, and weather. However, the RNN also has some obvious disadvantages. When the input time series data has a large dimension, the calculation efficiency of RNN is relatively low. Therefore, this study used a1D CNN module as a feature extractor to reduce the feature dimension and then used RNN for prediction. Specifically, the RNN module was a two-layer bidirectional Long short-term memory (bi-LSTM) [33], and the dimension of each layer was 64. As shown in Figure 5, bidirectional LSTM, as a special RNN, simultaneously models each sequence in both the forward and backward direction, which enables a richer representation of long-term dependencies of time series data, since each token encoding contains context information from the past and the future. Meanwhile, to avoid overfitting of the model, Dropout layers with coefficients of 0.2 and 0.5 were added between each layer of bidirectional LSTMs. Finally, the output vectors of the RNN were input to the softmax layer with 5 neurons to perform the heartbeat classification task.

4) DESCRIPTION OF TRANSFER LEARNING

Considering patient privacy and labor cost, it is usually difficult to obtain enough labeled ECG signal data. Therefore, limited by the scale of the training data, the proposed combined model still has a certain degree of deficiency in generalization. To further improve the model classification performance, we took advantage of transfer learning to improve the generalization and robustness of our deep learning model. We employed the 2017 PhysioNet/CinC Challenge dataset to pre-train our proposed model, and then transferred the learned knowledge to the heartbeat classification task. Specifically, the proposed transfer learning method included a pre-training phase and a parameter transfer phase. In the pre-training stage, we first adjusted the number of neuron nodes of the last fully connected layer of the combined network shown in Figure 1 to 4 to meet the needs of the pre-training task; then used the he_normal initialization method to set the initial weights of each layer of the combined network; finally used the 2017 PhysioNet/CinC Challenge dataset to train the combined network and saved the weights after training.

In the parameter migration phase, we first loaded the weights in the pre-trained model (except the weights in the last fully connected layer) into the combined network model to perform network parameter initialization. Then, we readjusted the number of neuron nodes in the last fully connected layer to 5 to meet the training needs of the target task. Finally, we used the training samples in the pre-processed MIT-BIH data set to train the model. After the model converged, we used the test set of MIT-BIH for testing. It was notable that this fine-tuning strategy in transfer learning did not increase the model parameter size while improving the model performance.

5) TRAINING METHOD

In the training process of our proposed neural network, we iteratively trained the network by randomly extracting small batches of data (batch_size) from the training data. The optimal setting of the learned weight parameters was obtained by minimizing loss functions that was determined by gradient descent algorithm. In this study, we used Keras (https://keras.io/) framework with Tensorflow as the backend to train and evaluate the proposed network. We used the combination of Adam optimizer [34] and cross-entropy loss function to find the optimal solution of the network. The cross-entropy loss function was used to calculate the difference between the prediction result of the sample and the real label, while the Adam optimizer comprehensively considered the first moment estimation (First Moment Estimation) and second moment estimation (Second Moment Estimation) of the gradient to update the adaptive weight parameters. After sufficient experiments, we set the three parameters of Adam optimizer's learning rate, beta-1, beta-2 to 0.001, 0.9, and 0.999, respectively. We set the batch_size to 250 and the number of iterations to 100 as well.

III. EXPERIMENTAL SETUP AND RESULTS

A. EXPERIMENTAL SETUP

1) EVALUATION

To evaluate the effectiveness of our model, we applied a train-test splitting scheme. For the MIT-BIH Arrhythmia Dataset, 800 ECG recordings were randomly selected as ''training ECG recordings'', and the actual training of the model was carried out on the training set. The remaining ECG recordings were set as ''test ECG recordings'', and the performance of the model was assessed on the test set. Accuracy (1) is the ratio of the correctly labeled subjects to the whole pool of subjects. Compared with classification accuracy, F-score (2) which is the current principal performance measure of a test's accuracy is a better metric when there are imbalanced classes [35]. F1-score that considered both the precision (3) and the recall (4) of the test was used over each class in our multiclass task. Precision is referred to as positive predictive value, while Recall is referred to as the true positive rate or

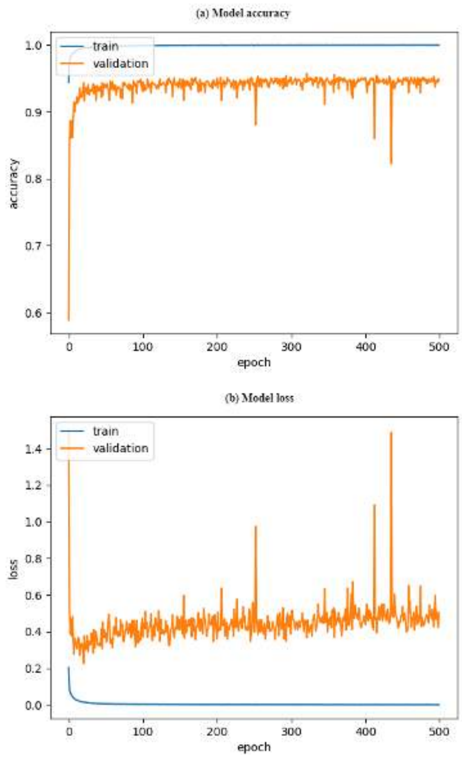

FIGURE 6. **Training and test curves for ECG beat classification. (a) accuracy over the training epochs; and (b) loss over the training epochs.**

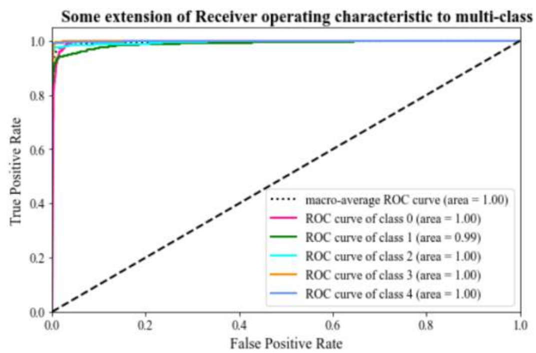

FIGURE 7. **ROC curves of our proposed network.**

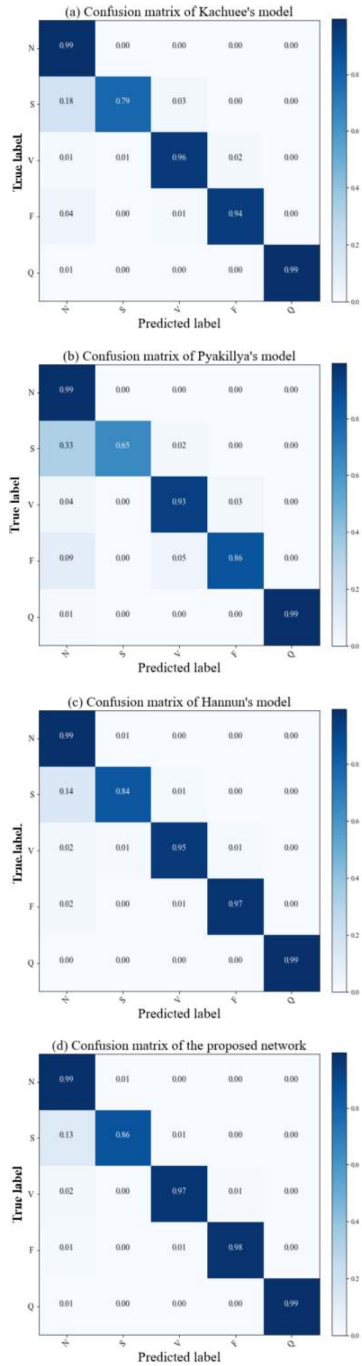

FIGURE 8. **Confusion matrix for ECG beat classification on the test set of MIT-BIH arrhythmia Dataset. (a), (b), (c) and (d) are confusion matrix of models of Kachuee *et al.* [36], Pyakillya *et al.* [37] and Hannun *et al.* [38], as well as our proposed network, respectively. Numbers inside blocks are number of samples classified in each category normalized by the total number of samples and rounded to two digits.**

sensitivity. The presented performance was comparable to performances reported in the literature. The area under the receiver operating characteristic curve (ROC) was also used to evaluate performance of our proposed network.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$F = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{2}$$

where

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

TP denotes the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives.

2) IMPLEMENTATION

The proposed network was implemented using the Keras (https://keras.io/) framework. Methods which did not involve

TABLE 2. Comparison of the proposed network and state-of-the-art methods.

Work	Method	Precision(%)	Recall(%)	F1-score(%)	Accuracy(%)
Kachuee et.al [36]	1D CNN	94.30	93.42	93.43	93.42
Pyakillya et.al[37]	1D CNN	91.35	88.52	88.58	88.52
Hannun et.al[38]	1D CNN	95.46	94.95	94.97	94.95
Our work	**1D CNN+ Bi-LSTM**	**96.34**	**95.90**	**95.92**	**95.90**

convolutional networks were coded in python. All experiments were performed under a Linux OS on a machine with CPU Intel Core i7-9700K @ 3.6 GHz, GPU NVIDIA GeForce GTX 1080, and 64 GB of RAM.

B. RESULTS

1) TRAINING AND TEST RESULTS

Our proposed network performed well for the beat epoch classification studies. When trained for 500 epochs, the classification accuracy rates were almost 100% for the training set of MIT-BIH arrhythmia Dataset, and the curve was relatively flat. While for the test set, the accuracies were between 90% and 95%, and the accuracy curve became relatively flat after 100 epochs (Figure 6a). Similarly, the loss curve of training set was almost 0 and very flat. While the test loss over epochs was between 0.3 and 0.5 (Figure 6b), indicating good generalization performance of the cross-entropy loss function. Figure 7 showed ROC curves of our network on the test dataset. The dashed (diagonal) line represented 0.5 ROC indicating a random performance. Our network achieved a high ROC of $\sim$1.

2) COMPARISON WITH THE STATE OF THE ART

Further comparison of confusion matrix (our proposed network vs. 3 state-of-the-art models) showed that our proposed network was able to make more accurate predictions and distinguish different classes with higher robustness (Figure 8a), while the models of Kachuee et.al, Pyakillya et.al and Hannun et.al had more miss-recognitions (Figure 8 b and c). The overall accuracy of 5-classes beat classification performed in previous section (3.2.1) presented an average accuracy of 95.9% (Figure 8d).

Moreover, we used 3 state-of-the-art models to classify the MIT-BIH arrhythmia database, and compared their precision, recall, and F1-scores with those of our proposed network. Compared with our proposed network presenting precision and recall performance of $\sim$96%, Kachuee's, Pyakillya's and Hannun's models had much lower precision and recall. Table 2 further demonstrated the higher average F1-score and accuracy in our proposed network. On the other hand, to further validate the superior performance of our network, we also collected results of classification work in the MIT-BIH arrhythmia database with other models.

IV. DISCUSSION AND CONCLUSION

Classification of ECG signals plays an important role in diagnoses of heart diseases. Applications of ECG signal classification models in detecting abnormality type and diagnosing a new patient are more precisely than manually. Developing and selecting the most appropriate classier is still a changeling problem.

Our proposed network showed better performance compared with the other 3 published models. As shown in table 2, the first row corresponded to a CNN-based model consisting of 1D CNN to extract ECG features and 5 residual blocks for classification. Each residual block involved 2 convolutional, 2 ReLu, a residual skip connection, and a pooling layer. The second CNN we tested was composed of 7 convolutional and 4 dense layers. All the CNN layers used 128 kernels, with the same size 5×5. While for the third CNN, it consists of 16 residual blocks with two convolutional layers per block. The 3 published deep learning algorithms were all one-dimensional CNN architectures. Although this network architecture had certain advantages in training efficiency, it mainly focused on the correlation with the feature space, and adopted the sliding window that could not capture long-distance features. Therefore, CNN was difficult to learn the long-term dependence of the features contained in the ECG sequence data in the time dimension. In addition, the hyperparameters such as convolution kernel size, number of convolution layers and pooling layers, as well as learning rate in the 3 published models would also affect the classification performance. While for our proposed classification method, we proposed a combined method of CNN and RNN. LSTMs is a type of RNN with a gated structure to learn long-term dependencies of sequence-based tasks. Bi-LSTMs are an extension of traditional LSTMs that train two instead of one LSTMs on the input sequence (one from past to future and one from future to past), which can understand context better, and thus can improve model performance on sequence classification problems. The combined method used the RNN to discriminate the features extracted by the CNN, so as to make better use of the advantages of CNN and LSTM (CNN has advantages in model size and training efficiency, LSTM has advantages in long-term dependence modeling of dynamic time series data). We also introduced a fine-tuning strategy in transfer learning, by transferring the knowledge learned in the 2017 PhysioNet/CinC Challenge dataset to the heartbeat classification task studied in this paper, in order to further improve the convergence speed of the combined model.

However, although the proposed network had better performance, there were still some limitations that could not be ignored and should be considered when interpreting the results. The MIT-BIH data set we used for training and testing had a considerable imbalance in category distribution. Although we adjusted the imbalance through the data amplification method, the data imbalance still has a certain impact on the generalization of the model performance. In addition, this study only studied one cardiovascular disease, i.e., arrhythmia, while the clinical signs of heart diseases were often complex and diverse. Thus, it will be necessary to continue to introduce more types of ECG data to expand the scope of the proposed network.

REFERENCES

[1] Y. Birnbaum, K. Nikus, P. Kligfield, M. Fiol, J. A. Barrabés, A. Sionis, O. Pahlm, J. G. Niebla, and A. B. de Luna, "The role of the ECG in diagnosis, risk estimation, and catheterization laboratory activation in patients with acute coronary syndromes: A consensus document," *Ann. Noninvasive Electrocardiol.*, vol. 19, no. 5, pp. 412–425, Sep. 2014.

[2] J. S. Steinberg *et al.*, "2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry," *Heart Rhythm*, vol. 14, no. 7, pp. e55–e96, Jul. 2017.

[3] M. Baumert, A. Porta, M. A. Vos, M. Malik, J. P. Couderc, P. Laguna, G. Piccirillo, G. L. Smith, L. G. Tereshchenko, and P. G. A. Volders, "QT interval variability in body surface ECG: Measurement, physiological basis, and clinical value: Position statement and consensus guidance endorsed by the European Heart Rhythm Association jointly with the ESC Working Group on Cardiac Cellular Electrophysiology," *Europace*, vol. 18, no. 6, pp. 44–925, Jun. 2016.

[4] J. Basu and A. Malhotra, "Interpreting the Athlete's ECG: Current state and future perspectives," *Current Treatment Options Cardiovascular Med.*, vol. 20, no. 12, p. 104, Nov. 2018.

[5] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ECG signals using machine learning techniques: A survey," in *Proc. Int. Conf. Adv. Comput. Eng. Appl.*, Mar. 2015, pp. 714–721.

[6] S. Goel, P. Tomar, and G. Kaur, "A fuzzy based approach for denoising of ECG signal using wavelet transform," *Int. J. Bio-Sci. Bio-Technol.*, vol. 8, no. 2, pp. 143–156, Apr. 2016.

[7] T. Barman, R. Ghongade, and A. Ratnaparkhi, "Rough set based segmentation and classification model for ECG," in *Proc. Conf. Adv. Signal Process. (CASP)*, Jun. 2016, pp. 18–23.

[8] A. Batra and V. Jawa, "Classification of arrhythmia using conjunction of machine learning algorithms and ECG diagnostic criteria," *Int. J. Biol. Biomed.*, vol. 1, no. 1, pp. 1–7, 2016.

[9] A. A. S. Raj, N. Dheetsith, S. S. Nair, and D. Ghosh, "Auto analysis of ECG signals using artificial neural network," in *Proc. Int. Conf. Sci. Eng. Manage. Res. (ICSEMR)*, Nov. 2014, pp. 1–4.

[10] Q. Zou, P. Xing, L. Wei, and B. Liu, "Gene2vec: Gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA," *RNA*, vol. 25, no. 2, pp. 205–218, Feb. 2019.

[11] F. Ren, C. Yang, Q. Qiu, N. Zeng, C. Cai, C. Hou, and Q. Zou, "Exploiting discriminative regions of brain slices based on 2D CNNs for Alzheimer's disease classification," *IEEE Access*, vol. 7, pp. 181423–181433, 2019.

[12] X. Gao, "Diagnosing abnormal electrocardiogram (ECG) via deep learning," in *Practical Applications of Electrocardiogram*, U. Lakshmanadoss, Ed. Rijeka, Croatia: InTech, 2019.

[13] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.

[14] Ö. Yıldım, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Comput. Biol. Med.*, vol. 102, pp. 411–420, Nov. 2018.

[15] P. Pławiak and U. R. Acharya, "Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals," *Neural Comput. Appl.*, pp. 1–25, Jan. 2019.

[16] U. B. Baloglu, M. Talo, O. Yildirim, R. S. Tan, and U. R. Acharya, "Classification of myocardial infarction with multi-lead ECG signals and deep CNN," *Pattern Recognit. Lett.*, vol. 122, pp. 23–30, May 2019.

[17] Y. Qiu, K. Huang, F. Xiao, and H. Shen, "Power-line interference suppression in electrocardiogram using recurrent neural networks," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–6.

[18] S. Singh, S. K. Pandey, U. Pawar, and R. R. Janghel, "Classification of ECG arrhythmia using recurrent neural networks," *Procedia Comput. Sci.*, vol. 132, pp. 1290–1297, Jan. 2018.

[19] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," *IEEE Access*, vol. 6, pp. 36188–36197, 2018.

[20] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network," *Sci. Rep.*, vol. 9, no. 1, p. 6734, May 2019.

[21] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 104964.

[22] M. Salem, S. Taheri, and J.-S. Yuan, "ECG arrhythmia classification using transfer learning from 2- dimensional deep CNN features," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2018, pp. 1–4.

[23] R. Xiao, Y. Xu, M. M. Pelter, D. W. Mortara, and X. Hu, "A deep learning approach to examine ischemic ST changes in ambulatory ECG recordings," *AMIA Summits Transl. Sci. Proc.*, vol. 2017, pp. 256–262, May 2018.

[24] Y. Wu, F. Yang, Y. Liu, X. Zha, and S. Yuan, "A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification," in *Proc. Comput. Vis. Pattern Recognit.*, 2018.

[25] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, p. E215-20, Jun. 2000.

[26] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[27] H. M. Rai, A. Trivedi, and S. Shukla, "ECG signal processing for abnormalities detection using multi-resolution wavelet transform and artificial neural network classifier," *Measurement*, vol. 46, no. 9, pp. 3238–3246, Nov. 2013.

[28] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, Sep. 2013.

[29] S. Banerjee and M. Mitra, "Application of cross wavelet transform for ECG pattern analysis and classification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 326–333, Feb. 2014.

[30] T. Liu, A. Abd-Elrahman, J. Morton, and V. L. Wilhelm, "Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system," *GIScience Remote Sens.*, vol. 55, no. 2, pp. 243–264, Mar. 2018.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, p. 8.

[32] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.

[33] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," *CoRR*, vol. abs/1801.02143, Jan. 2018.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[35] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 4304, A. Sattar and B. Kang, Eds. Berlin, Germany: Springer, 2006.

[36] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 443–444.

[37] B. Pyakillya, N. Kazachenko, and N. Mikhailovsky, "Deep learning for ECG classification," *J. Phys., Conf. Ser.*, vol. 913, Oct. 2017, Art. no. 012004.

[38] 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 Med.*, vol. 25, no. 1, pp. 65–69, Jan. 2019.

XUE XU received the Ph.D. degree from Northwest A&F University, China, where she focused on designing and constructing machine learning models to explore disease mechanisms. She is currently a Postdoctoral Research Fellow with the Marcus Institute for Aging Research, Harvard Medical School. She is also focusing on biological signal data, and is combining the computational analysis and the wet experiments. She is also particularly interested in age-related diseases.

SOHYUN JEONG received her Ph.D. degree from the College of Pharmacy, Seoul National University, Seoul Korea. She is currently an Assistant Scientist I with the Marcus Institute for Aging Research, Hebrew SeniorLife and Harvard Medical School. Her major research areas are pharmacogenomics, pharmacoepidemiology and human genetics utilizing population-based big data, GWAS, and microbiome data.

JIANQIANG LI (Member, IEEE) is currently a Professor with the College of Computer Science and Software Engineering, Shenzhen University, China. His research interests are the Internet of Things, cyber physics systems, mobile computing, and artificial intelligence.