Biomedical Signal Processing and Control 52 (2019) 77–83

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Automatic staging model of heart failure based on deep learning

Dengao Li*, Xuemei Li, Jumin Zhao, Xiaohong Bai

School of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

ARTICLE INFO

Article history:
Received 18 August 2018
Received in revised form 14 February 2019
Accepted 16 March 2019

Keywords:
Heart failure
Staging model
Deep learning
Deep CNN-RNN model

ABSTRACT

Heart failure (HF) is a disease that is harmful to human health. Recent advances in machine learning yielded new techniques to train deep neural networks, which resulted in highly successful applications in many pattern recognition tasks such as object detection and speech recognition. To improve the diagnostic accuracy of HF staging, this study evaluates the performance of deep learning-based models on combined features for its categorization. We proposed a novel deep convolutional neural network-Recurrent neural network (CNN-RNN) model for automatic staging of heart failure diseases in real-time and dynamically. We employed the data segmentation and data augmentation pre-processing dataset to make the classification performance of the proposed architecture better. Specifically, this paper use convolutional neural network (CNN) as a feature extractor instead of training the entire network to extract the characteristics of the electrocardiogram (ECG) signals and form a feature set. We combine the above feature set with other clinical features, feed the combined features to RNN for classification, and finally obtain 5 classification results. Experiments shows that the CNN-RNN model proposed in this paper achieved an accuracy of 97.6%, the sensitivity of 96.3%, specificity of 97.4% and proportion of 97.1% for two seconds of ECG segments. We obtained an accuracy, sensitivity, specificity and proportion of 96.2%, 96.9%, 95.7%, and 94.3% respectively for five seconds of ECG duration. The model can be used as an aid to help clinicians confirm their diagnosis.

1. Introduction

Heart failure (HF) is a serious or terminal stage of various cardiovascular diseases (CVDs), with high disability and mortality, and 5-year survival rate is comparable to malignant tumors [1]. Since 1980, the American College of Cardiology (ACC) and American Heart Association (AHA) have translated scientific evidence into clinical practice guidelines (guidelines) with recommendations to improve cardiovascular health [2]. These guidelines, which are based on systematic methods to evaluate and classify evidence, provide a cornerstone for quality cardiovascular care. Based on clinical features, patients from only cardiovascular risk factors to heart failure, the development of refractory end-stage heart failure, and the final process of death is divided into four stages: stage A, stage B, stage C and stage D, and the severity of the representative from A to D increases in turn. The ACC/AHA guideline is obtained worldwide universal recognition [3]. Heart failure is a process of continuous progress. Patients can only stay in a certain period or progress forward and cannot be reversed. According to the staging of heart failure, there are two important periods of intervention: prevention of progression from stage A to stage B, and prevention of progression of stage B to stage C [4]. Early detection of heart failure and the correct diagnosis of heart failure staging is the basis for ideal therapeutic results, but due to the lack of simple and effective staging diagnosis model of heart failure, it is difficult to diagnose heart failure clinically, resulting in the diagnosis rate and control rate of heart failure relatively low [5]. Therefore, it is very important to establish a real-time and dynamic heart failure automatic staging model.

Recently the authors of [6] proposed a monitoring system for clinical management of HF. Random Forests algorithm was employed based on its performance. It was evaluated in terms of accuracy, sensitivity and specificity for each class versus all the other classes in a 10-fold cross validation. The obtained accuracy was 81.3%, while the sensitivity and specificity were 87% and 95%, respectively for class 3 (severe HF). Class 1 (mild HF) was identified with 75% sensitivity and 84% specificity and class 2 (moderate HF) was identified with 67% sensitivity and 80% specificity. Shahbazi et al. 2015 [7] exploited long-term HRV measures to estimate the severity of HF and more specifically to classify patients to low risk and high risk. Generalized Discriminant Analysis was applied for reducing the number of features, as well as to overcome over-

* Corresponding author.
E-mail address: lidengao@tyut.edu.cn (D. Li).

https://doi.org/10.1016/j.bspc.2019.03.009
1746-8094/

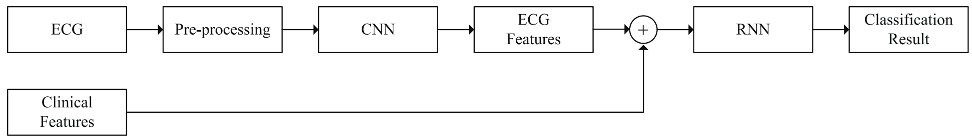


Fig. 1. The flow diagram of this paper.

lapping of the samples of two classes in the feature space. The selected features were given as input to a k-NN classifier providing classification accuracy 97.43% in the case. Yang et al. 2010 [8] proposed a scoring model allowing classification of a subject to three groups: health group (without cardiac dysfunction), HF-prone group (asymptomatic stages of cardiac dysfunction) and HF group (symptomatic stages of cardiac dysfunction). SVM was employed and the total accuracy was 74.40%. The accuracy for each one of the three groups was 78.79% for healthy group, 87.50% for HF-prone group and 65.85% for the HF group.

Despite the fact that a lot of research has been done in ECG assessment, we believe that HF automatic stage still needs to be addressed. For this reason, we designed a novel model and use lots of data for training and testing our algorithms. The clinical relevance is also assessed through expert assessment and compared against the baseline provided by an automatic diagnostics system routinely used in clinical practice.

In this work, we propose a novel deep CNN-RNN model for the automatic staging of heart failure diseases. The deep learning model consists of two parts: the CNN is used as a feature extractor to extract the characteristics of the ECG signals to form a feature set; the above ECG features are combined with other clinical features of the patient, then input into the RNN, and finally the classification results are obtained. The deep learning model proposed in this paper does not require denoising the signal and does not require feature extraction. Instead of using CNN classification alone, we combine CNN with RNN to propose a new classification model. Experimental results on a realworld dataset demonstrate the performance of the proposed staging model. The deep CNN-RNN model presented in this paper is used to automatically stage heart failure diseases in real time and dynamically. It improves the diagnostic accuracy of HF staging and can help clinicians confirm their diagnosis and reduce misdiagnosis. The flow diagram of this paper is shown in Fig. 1.

The rest of the paper is organized as follows. In Section 2, we introduced the methodology, including dataset processing and the proposed modeling architecture. Section 3 introduces the results, we evaluate it using real-world data and discuss the work. We conclude our work and highlight future research directions in Section 4.

2. Methodology

2.1. Dataset

The dataset used in this paper is from the chest pain centers (CPCs) of Shanxi Academy of Medical Sciences. Our study is a retrospective study. The valid data we collected included 573 patients (aged at least 18 years old) from January 2013 to December 2017, including healthy person and heart failure patients. The dataset consists of ECG signals, as well as gender, age, coronary heart disease, hypertension, history of diabetes, primary percutaneous coronary intervention (PCI). Data other than ECG signals are called clinical data.

Recent updates of AHA/ACC guidelines about HF focus on diagnosis and management of HF in adults. The current methods for clinical assessment of HF subdivide patients with HF or with high risk for HF in four stages: A, B, C and D. In the first two stages (A and B), the patients are asymptomatic, whereas in the last two stages, they have a clinically manifested HF (stage C) which becomes refractory to therapy (stage D). The current guidelines modulate therapeutic interventions on the basis of this classification [2].

According the guidelines, all ECG signals should be classifed into five classes: Normal (N), stage A (A), stage B (B), stage C (C) and stage D (D). The five classes of ECG signals are shown in Fig. 2.

In this work, we have used lead II ECG signals. The overview of the data used in this study is shown in Table 1.

Table 1
Overview of the data used in this study.

Dataset	Number	Proportion
Set N	172	30.0%
Set A	84	14.7%
Set B	156	27.2%
Set C	105	18.3%
Set D	56	9.8%
Total	573	100%

Table 2
Summary of the segments.

Dataset	Number
Set I	17,190
Set II	6876
Total	24,066

2.2. Pre-processing

2.2.1. Signal segmentation

Given the ECG signals, they should be divided into separate waveforms firstly. The ECG signals we collect are sampled at a frequency of 250 Hz, and each ECG record has a duration of 1 min. We divide them into two data sets: set I and set II. Set I has a duration of 2 s, and set II has a duration of 5 s. After segmentation, each ECG signal is regularized with Z score normalization, standard deviation of 1, and zero mean before inputting into the network. Finally, we obtain 17190 2 s ECG segments and 6876 5 s ECG segments (Table 2) and feed them separately to the deep learning model for training and testing.

2.2.2. Data augmentation

The ECG record is divided into five categories, corresponding to one healthy state normal (N) and four staging (both left heart failure): A, B, C and D. The number of five ECG signals is not equal. To overcome the class imbalance problem [9], we generated synthetic data as follows: (1) Change the standard deviation and mean of Z-score calculated from the original normalized ECG signals by 10%. (2) Produce a random Gaussian noise signal with a mean of 0 and standard deviation of 0.05. (3) Combine the above two signals to produce the new signal. In order to make the effect of deep neural network training better, the segments in Set N remain unchanged because they are the most abundant, and the number of segments of the remaining type is increased to match the N class. The number of segments increases the number of A, B, C, and D to be the same as N.

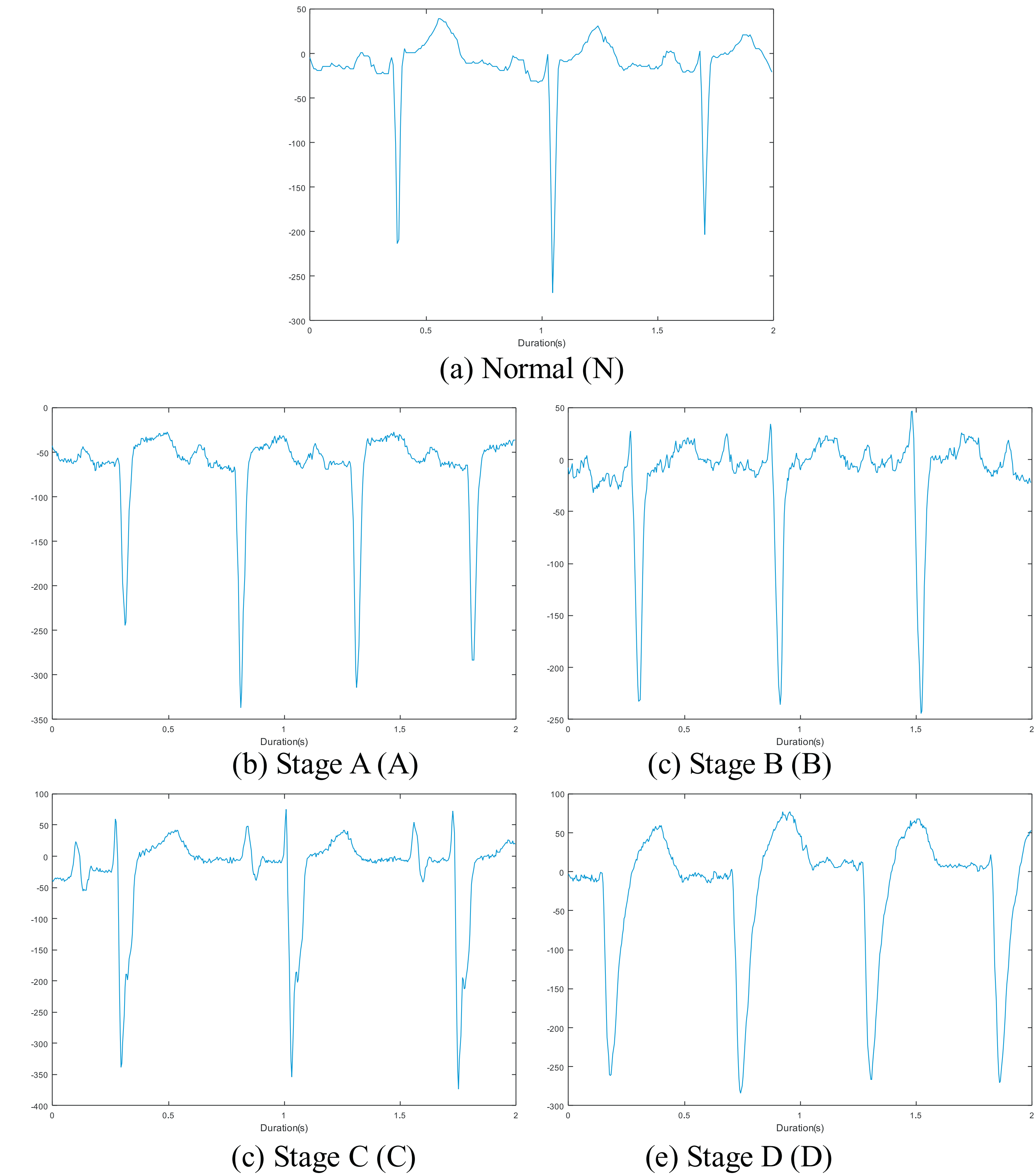


Fig. 2. Illustrations for ECG signals of different classes.

Thus, the 2 s ECG segments are both 5160, and the 5 s ECG segments are both 2064, they forming a balanced dataset (Table 3). After the enhancement, the total number of Set I segments including N, A, B, C, and D increases to 5160*5 = 25800, and the total number of Set II segments increases to 2064*5 = 10,320. We augment the clinical data of each patient to form a balanced feature set similarly.

In addition, this paper intends to use the convolutional neural network to process ECG signals. We note that CNN is a special network that eliminates the need for pre-processing and separate feature extraction technique [10]. Therefore, our algorithm does not require denoising. It can process ECG signals without filtering any noise present in the ECG signal.

Table 3
Summary with the breakdown of the 5 classes of ECG signals.

Class	2 s ECG Segments	5 s ECG Segments
Set N	5160	2064
Set A	2520	1008
Set B	4680	1872
Set C	3150	1260
Set D	1680	672
Total	17,190	6876

2.3. Deep learning model

2.3.1. 1-Dimensional Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep learning model form, and CNN is one of the most commonly used artificial neural networks [11]. It was first proposed by Fukushima in 1980 [12], and improved by LeCun et al later [13]. CNN is a back propagation neural network consisting of many hidden layers and parameters. And its artificial neurons can respond to surrounding units in a part of the coverage. CNN has been successfully applied to many applications such as target detection [14,15], face recognition [16,17] and image classification [18,19]. It is also used as an automated diagnostic tool in the medical field as a computer-aided diagnostic system [20–24]. Due to the characteristics of the ECG itself, that is, the one-dimensionality different from the two-dimensional image, this paper uses 1-dimensional CNN to achieve better results.

The structure of the CNN includes convolutional layer, pooling layer, and fully connected layer. In particular, the CNN network is used to automatically learn and acquire different features from the input ECG signals without the need for separate pre-processing and feature extraction steps. Therefore, it can help reduce the burden during training and select the best feature extraction technique for automatic staging of heart failure. In addition, if we can implement the learning-based hidden layer fitting learning by learning the structure of the data, it is possible to obtain better performance. Therefore, we consider CNN as a feature extractor instead of training the entire network to extract the characteristics of ECG signals and form a feature set.

This paper uses an 8-layer deep CNN, including three convolution layers, three maxpooling layers and two fully-connected layers. Behind each convolutional layer is a pooling layer. The role of the pooling layer is to reduce the size of the feature map, and finally the full connection layer outputs 20 features. The specific structure is shown in Fig. 3.

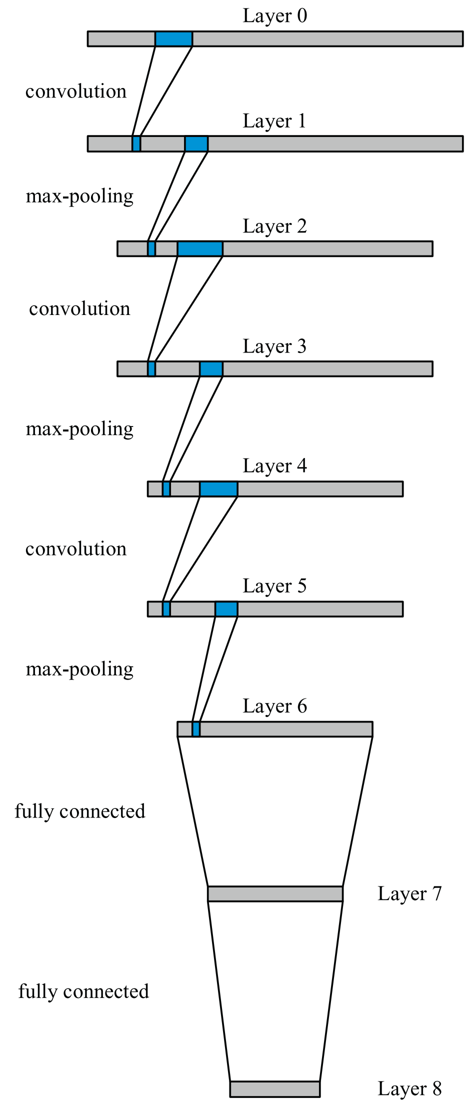


Fig. 3. 1-dimensional Convolutional Neural Network.

2.3.2. Recurrent neural network

The working mechanism of recurrent neural network (RNN) is very similar to that of animal brain [25]. It consists of input layer, hidden layer and output layer. In particular, the result of the hidden layer is related to the input of the current layer and the output of the previous layer. Hochreiter et al. proposed the LSTM model in 1997 (Fig. 4), which is a special RNN model [26]. The layers of the LSTM are added to the valve nodes based on the original RNN network, which is beneficial to overcome the problem of RNN long-term memory calculation.

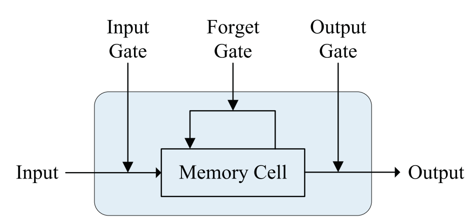


Fig. 4. Internal structure of LSTM memory block.

Considering the network simplicity and the classifcation performance, we propose a four-layer RNN to classify the combined dataset, as shown in Fig. 5. The frst two layers are LSTM layers, including 50 and 150 memory blocks, respectively. The last two layers are fully-connected (FC) layers. In particular, the third layer includes 20 neurons, while the fourth layer (the output layer) has 4 neurons according to the number of classes.

2.3.3. CNN-RNN architecture

The CNN part of our proposed CNN-RNN model extracts features from ECG signals, and the RNN portion simulates ECG signal/clinical feature relationships and feature dependencies. We input the previously extracted ECG features along with the clinical data into the RNN and finally output the classification results.

We initialize the model weights using Xavier initialization.

Each convolution layer is convolved with their respective kernel sizes using Eq. (1)

$$x_n = \sum_{k=0}^{N-1} y_k f_{n-k} \quad (1)$$

Where y, f and N represent the signal in y, the number of filters and elements, and x represents the output vector.

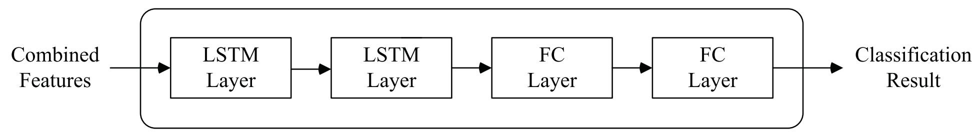


Fig. 5. The RNN proposed in this paper.

We use Rectified Linear Units (ReLU) to facilitate learning of non-linearly separable problems.

$$f(x) = \max(0, x) \tag{2}$$

ReLU can also be written as the Eq. (3)

$$f(x) = \begin{cases} x, & if\, x \geq 0 \\ 0, & if\, x < 0 \end{cases} \tag{3}$$

The CNN-RNN model proposed in this paper is trained using backpropagation with a sample size of 10. Network loss is assessed using cross-entropy loss (Eq. (4)).

$$C(y, a) = -\frac{1}{n}\sum_{n} y\log(a) + (1-y)\log(1-a) \tag{4}$$

The hyperparameters are set as follows: regularization (λ) is 0.2, learning rate is $3^{*}10^{-3}$, and momentum parameter is set to 0.7. Such settings prevent over-fitting of data, assist in data convergence, and control the speed of learning during training. Further, the deviation and the weight are updated by the Eqs. (5) and (6).

$$\Delta W_l(t+1) = -\frac{x_\lambda}{r}W_l - \frac{x}{n}\frac{\partial C}{\partial W_l} + mW_l(t) \tag{5}$$

$$\Delta B_l(t+1) = -\frac{x}{n}\frac{\partial C}{\partial B_l} + m\Delta B_l(t) \tag{6}$$

where W, B, l, λ, x, n, m, t, and C denotes the weight, bias, layer number, regularization parameter, learning rate, total number of training samples, momentum, updating step, and loss function respectively.

We set the initial learning rate to 0.003, and every 300 epochs, the learning rate is reduced to 1/10 of the initial learning rate until the neural network converges. The training method uses a ten-fold cross-validation [27] approach to randomly divide the total data set including set I and set II into 10 equal parts. The training process of the CNN-RNN model includes 10 iterations. In each iteration, 9 parts of the data were used for training and 1 parts of the data was used for testing, so iteratively performed 10 times. In each fold, performance is assessed, ie specificity, sensitivity and accuracy. The model in each iteration is based on the final model in the previous iteration. All ten times the average gives the overall performance of the system. Finally, the CNN-RNN model output 5 classification results.

3. Results

We used four standard metrics to evaluate the classification performance: classification accuracy (Acc), sensitivity (Se), specificity (Sp), and positive predictivity (Pp). Using true positive (TP), false positive (FP), true negative (TN), false positive (FN), Acc, Se, Sp, Pp are defined as follows.

Acc is the ratio between the number of correctly classifed patterns and the total number of patterns.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Table 4
Classification results of different duration.

Dataset	Feature Combination				No Feature Combination			
	Acc	Se	Sp	Pp	Acc	Se	Sp	Pp
Set I	**97.6%**	96.3%	**97.4%**	**97.1%**	92.9%	93.1%	89.4%	90.5%
Set II	96.2%	**96.9%**	95.7%	94.3%	90.5%	89.6%	92.8%	88.1%

Se is the proportion of the number of correctly detected events in all events.

$$Se = \frac{TP}{TP + FN} \tag{8}$$

Sp is the ratio of correctly classifed nonevents among nonevents.

$$Sp = \frac{TN}{TN + FP} \tag{9}$$

Pp is the proportion of correctly detected events in all detected events.

$$Pp = \frac{TP}{TP + FP} \tag{10}$$

We performed two sets of experiments on the two sets of data. The first set of experiments used combined data of ECG signals and other clinical features. The second set of experiments did not use combined data and only used ECG signals alone for experiments. The experimental results are shown in Table 4.

We found that Set I with the combined features had the highest accuracy of 97.6%, followed by Set II with the combined features, with an accuracy of 96.2%. The accuracy of Set I and Set II without the combined features were 92.9% and 90.5%, respectively. Combining the three indicators of Se, Sp and Pp, Set I with the combined features has the best classification effect.

In addition, we compared the results of four experiments with two sets of experimental data. Fig. 6 shows the ROCs obtained with the different sets of contrast duration. We found that the ECG signal with the duration of 2 s combined with clinical features had the highest AUC value of 0.851. This is followed by an ECG signal that is 5 s in duration and incorporates clinical features. The lowest AUC value is ECG data that is 5 s in length but does not incorporate clinical features.

The above experimental results show that the classification effect of ECG signal combined with other clinical features is better than that of ECG alone. However, we do not know that the relationship between those features in clinical features and heart failure staging is higher. Therefore, we plan to explore an algorithm for calculating the weights of different features and heart failure staging, and obtain reasonable results to optimize our proposed deep automatic staging model. In addition, since the causes of heart failure are complex and diverse, there are many indicators to be used for staging diagnosis. This article only selects some of them, so there are inevitable limitations. In the future we plan to add more features to the training set and perform more comprehensive and precise model.

We compared the deep CNN-RNN model that proposed in this paper with other methods and found that our model is superior in terms of performance (Table 5). In the control experiment, we all used 2 s ECG segmentation and we combined ECG signals and clinical data.

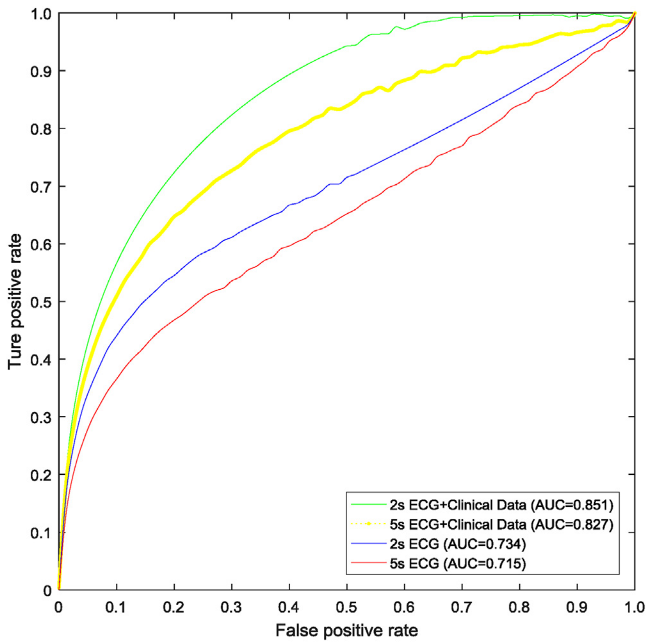


Fig. 6. ROC curves of the contrast experiment.

Table 5
Summary of methods comparison using the same database.

Method	Acc	Se	Sp	Pp
CNN-RNN	**97.6%**	**96.3%**	**97.4%**	**97.1%**
MLP	93.3%	85.7%	84.4%	79.8%
RF	82.1%	83.4%	81.7%	80.5%
CART	72.3%	76.6%	78.8%	77.9%
SVM	65.9%	73.3%	61.2%	66.8%

(CNN-RNN: Convolutional Neural Network-Recurrent Neural Network, MLP: Multi-layer Perceptron, RF: Random Forest, CART: Classification and Regression Tree, SVM: Support Vector Machine).

4. Discussion

We analyzed the data used in the experiments in this paper and visually analyzed the values of ECG signals (Fig. 7). From the box-plots we can see that the normal ECG signal and the heart failure ECG signal show a significant difference, but the difference between the four categories heart failure ECG signals (A, B, C and D) is not obvious in the figure, we need to use the proposed method in this paper re-analyzes them.

In this work, a combination of CNN with RNN in a single network model was proposed so that the overall performance of the model can be optimized through training. Both CNN and RNN models learn different functions and hence, merging of these two nets has yielded higher diagnostic accuracy. Furthermore, the data augmentation in this work has also helped improve the robustness of the proposed algorithm as it can generate more ECG segments with slight variations to train the model.

The main benefits of our proposed model are as follows:

(1) Feature extraction, selection and classification procedures are combined in a single deep CNN-RNN structure.
(2) Denoising is not required.
(3) Ten-fold cross validation ensures the results are reliable and robust.

The main drawbacks of our proposed model are as follows:

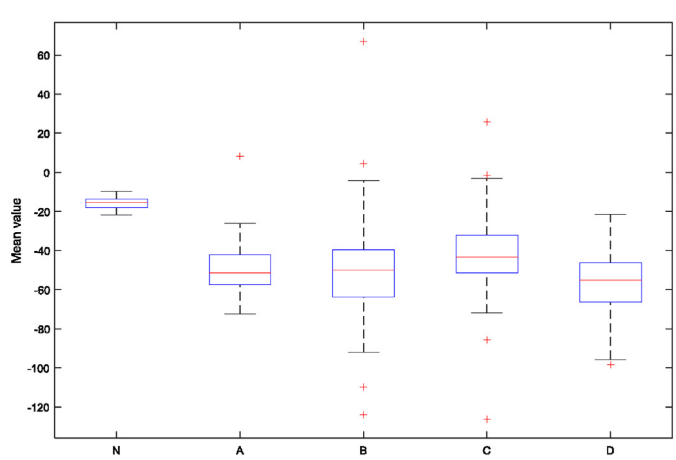
(1) Requires a lot of data for training.
(2) Takes more time to train the data.

5. Conclusion

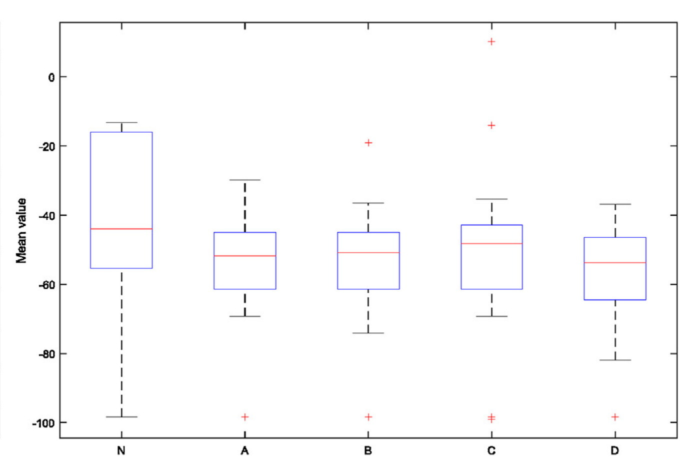
In this paper, we proposed a novel CNN-RNN model for automatic staging of heart failure diseases based on deep learning. We uses CNN as a feature extractor and employ RNN for classification. Feature extraction, selection and classification procedures are combined in a single deep CNN-RNN structure. Experiments prove that the CNN-RNN model proposed in this paper achieved good robust and generalization performance on real dataset. Future work will include incorporating expert knowledge into our framework and expanding our approach to additional health care applications.

Acknowledgments

The paper supported by The General Object of National Natural Science Foundation (61772358): Research on the key technology of BDS precision positioning in complex landform; Project supported



(a) Two-second ECG segments



(b) Five-second ECG segments

Fig. 7. Mean value of different intervals of ECG signals.

by International Cooperation Project of Shanxi Province (Grant No. 201603D421014): Three-Dimensional Reconstruction Research of Quantitative Multiple Sclerosis Demyelination; International Cooperation Project of Shanxi Province (Grant No. 201603D421012): Research on the key technology of GNSS area strengthen information extraction based on crowd sensing.

References

[1] H. Huang, B. Huang, Y. Li, et al., Uric acid and risk of heart failure: a systematic review and meta-analysis, Eur. J. Heart Fail. 16 (1) (2014) 15–24.

[2] C.W. Yancy, M. Jessup, B. Bozkurt, et al., ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure, J. Am. Coll. Cardiol. 68 (13) (2017) 1476–1488, 2017.

[3] B. Bozkurt, What is new in heart failure management in 2017? Update on ACC/AHA heart failure guidelines, Curr. Cardiol. Rep. 20 (6) (2018) 39.

[4] D.K. Jin, K. Shido, H.G. Kopp, et al., Guidelines for the diagnosis and treatment of chronic heart failure, in: Joseph Massie on the Natural Rate of Interest, 1750, Lord Baltimore press, 2005, pp. 35–53.

[5] M. Alfarano, P. Severino, M. Pucci, et al., A TNM-like staging system for risk stratification in heart failure patients, J. Am. Coll. Cardiol. 71 (11) (2018), A822.

[6] G. Guidi, L. Pollonini, C.C. Dacso, et al., A multi-layer monitoring system for clinical management of Congestive Heart Failure, BMC Med. Inform. Decis. Mak. 15 (Suppl. 3) (2015), S5-S5.

[7] F. Shahbazi, B.M. Asl, Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability, Comput. Methods Programs Biomed. 122 (2) (2015) 191–198.

[8] G. Yang, Y. Ren, Q. Pan, et al., A heart failure diagnosis model based on support vector machine, International Conference on Biomedical Engineering & Informatics (2010).

[9] R. Longadge, S. Dongre, Class imbalance problem in data mining review, Int. J. Comput. Sci. Netw. 2 (1) (2013).

[10] Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and time series, Handbook Brain Theory Neural Netw. 1995 (10) (1995) 3361.

[11] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.

[12] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (4) (1980) 193–202.

[13] Y. Lecun, L. Bottou, Y. Bengio, et al., Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[14] X. Jiang, C. Wang, Q. Fu, Development and application of deep convolutional neural network in target detection, in: Advances in Materials, Machinery, Electronics Ii: Proceedings of the, International Conference on Advances in Materials, Machinery, Electronics, 2018, 040036.

[15] D.U. Lan, B. Liu, W. Yan, et al., Target detection method based on convolutional neural network for SAR image, J. Electron. Inf. Technol. 38 (12) (2016) 3018–3025.

[16] Y. Xi, X. Liu, Multi-task convolutional neural network for pose-invariant face recognition, IEEE Trans. Image Process. PP (99) (2017), 1-1.

[17] N. Zhuang, Y. Yan, S. Chen, et al., Multi-label learning based deep transfer neural network for facial attribute classification, Pattern Recognit. (2018) 80.

[18] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, International Conference on Neural Information Processing Systems (2012) 1097–1105.

[19] P. Shamsolmoali, M. Zareapoor, J. Yang, Convolutional neural network in network (CNNiN): hyperspectral image classification and dimensionality reduction, IET Image Process. 13 (2) (2019) 246–253.

[20] S. Hoochang, H.R. Roth, M. Gao, et al., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285.

[21] R. Rasti, M. Teshnehlab, S.L. Phung, Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks, Pattern Recognit. (2017) 72.

[22] J. Ma, F. Wu, J. Zhu, et al., A pre-trained convolutional neural network based method for thyroid nodule diagnosis, Ultrasonics 73 (2016) 221.

[23] Y. Komeda, H. Handa, T. Watanabe, et al., Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience, Oncology 93 (Suppl 1(1)) (2017) 30–34.

[24] T. Kooi, G. Litjens, B.V. Ginneken, et al., Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2016) 303.

[25] F.J. Pineda, Generalization of back-propagation to recurrent neural networks, Phys. Rev. Lett. 59 (19) (1987) 2229–2232.

[26] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[27] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second edition, 2001.