Application of Deep Neural Network for Congestive Heart Failure Detection Using ECG Signals

Yue Zhang\textsuperscript{1, a} and Ming Xia\textsuperscript{1, b}

\textsuperscript{1}Division of Information Science and Technology at Shenzhen, Tsinghua University, China

\textsuperscript{a} zhangyue@mails.tsinghua.edu.cn; \textsuperscript{b} 1034310291@qq.com

\textbf{Abstract.} Congestive heart failure (CHF) is a common serious heart disease that requires a number of clinical examinations to diagnose, which are costly and time-consuming. Electrocardiogram (ECG) is widely used in the diagnosis of various cardiovascular diseases due to its advantages of non-invasive, convenient and cheap, so the automatic CHF detection algorithm based on ECG signals can overcome the above shortcomings and has great application prospects. In this paper, inspired by the idea of DenseNet in computer vision, we refined it to be applicable to CHF detection task, thus improving the diagnosis accuracy of the model. Secondly, to improve the robustness of the algorithm, we built a CHF database on PhysioBank, which contained more diverse data compared with similar studies, and conducted experiments on the built database. Finally, we presented an evaluation method based on the “inter-patient” pattern to evaluate the performance of the method more objectively. The results show that our algorithm can efficiently detect CHF with accuracy, sensitivity and specificity up to 94.97\%, 89.38\% and 99.50\%, respectively. The algorithm proposed in this study can provide reliable references for doctors, and can be used in portable devices to realize real-time monitoring for patients.

1. Introduction

Congestive heart failure is a common heart disease in which the heart can’t deliver enough oxygen-rich blood to other tissues and organs due to structural or functional disorders of the cardiac. At the same time, the insufficient pumping function of the ventricle causes blood and body fluid of patient to flow back to the lungs and body, leading to respiratory distress and systemic swelling\cite{1}. According to statistics, about 26 million people worldwide have been diagnosed with heart failure, which has become a major cause of global mortality and morbidity\cite{2}. Electrocardiogram is widely applied in the clinical diagnosis of heart diseases by capturing the weak potential changes on the body surface visa electrodes placed on the body surface, so as to reflect the status of the heart. Therefore, it is of great practical value and market potential to study the diagnosis method for CHF using ECG signals.

Algorithms for detecting CHF using ECG signals have been studied in the past decades and have achieved good performance. They mainly use feature extraction-based machine learning methods, where the extracted features include morphological features, features in time and frequency domain, etc., then the extracted features are selected and entered into the classifier for diagnosis. In order to get better classification results, a combination of multiple feature parameters is needed to train the model. For example, Sharma et al. \cite{3} used eigenvalue decomposition of Hankel matrix to extract the lowest and highest frequency components of the signal, so as to obtain the mean and standard deviation in
time domain, K-NN entropy, correlation entropy and the average frequency as features. The algorithm adopted LS-SVM to detect CHF. In [4], Acharya et al. applied empirical mode decomposition to decompose signals to obtain IMFs, from which they extracted 11 entropy features and 2 complexity features. After features selection, SVM was used for classification, and the accuracy of 97.64% was obtained. It can be seen that the main work of these methods lies in the manual features extraction step, which has a crucial impact on the results. With the continuous success of deep learning in other fields, many researchers have introduced it into ECG recognition area. In [5], Pranav et al. trained a 34-layer CNN to detect arrhythmia using ECG signals, and the diagnosis results were higher than that of human doctors, demonstrating the application prospects of deep learning models in medical fields. Therefore, some researchers started to use deep learning algorithms to detect CHF. For example, Chen et al. [6] used a sparse-auto-encoder to extract unsupervised features of the original signals, and then applied fully connected neural networks with different combinations of hidden nodes to detect CHF with a model accuracy of 72.41%. Wang et al. [7] designed a LSTM network based on the short RR intervals to detect congestive heart failure, demonstrating the ability of RR interval to diagnose CHF. In [8], Acharya et al. trained an 11-layer convolutional neural network to automatically identify heart failure. The model achieved 98.97% accuracy on the dataset divided based on the "intra-patient" mode, indicating the effectiveness of using deep learning for CHF detection.

In summary, most of the CHF detection methods are machine learning algorithms based on feature extraction, which heavily rely on human experience and require complex feature extraction and feature selection work. However, the CHF detection algorithm using deep learning is rarely studied at present, whose performance is not good enough. In addition, almost all CHF detection methods currently use the "intra-patient" pattern to divide data, resulting in ECG data from one person appears in training set and testing set at the same time, which is somewhat similar to data leakage. High accuracy is easy to achieve, but the model's generalization ability drops rapidly. Therefore, the evaluation method also needs to be changed to match the practical applications.

In this paper, we presented a novel detection algorithm to detect congestive heart failure. Firstly, in order to meet the demand for data volume in the training stage, we built a database on PhysioBank, which contained more diverse data compared with similar studies. Secondly, inspired by the convolutional neural network DenseNet, we applied the idea of this model to CHF detection task to improve the diagnosis accuracy. Finally, we divided the data based on the "inter-patient" pattern. One person's data only appeared in the training set or the testing set, which was more in line with the actual application scenarios, thus giving objective evaluation indexes of the algorithm.

The arrangements of the article are as follows: the methodologies and related theories are presented in the following part, mainly including building database, data pre-processing, network structure, etc. The third part presents the parameter setting, evaluation method, experimental result and discussion. Section IV summarizes the presented method in this article.

2. Methods

2.1. Data description

In this paper, we built a CHF diagnosis database on the public physiological database PhysioBank [9], and the built database contained twice as much data as the current CHF diagnosis algorithms using ECG segments(not HRV signals). To create this CHF database, we look through every ECG record in all ECG databases on PhysioBank and selected the records diagnosed with heart failure from tens of thousands of records to form the CHF dataset. It was a huge, tedious but meaningful work. Ultimately, the database contained 33 patients from 6 databases and 58 healthy subjects (control group) from 2 databases. The detailed information is shown in Table 1.

2.2. Pre-processing process

The ECG signal of human body is a weak physiological signal which is extremely vulnerable to the internal and external environment during the acquisition process, and collected signals are often
accompanied by strong noises that drown out the useful characteristics. Failure to remove noises will cause the algorithm to learn some noise patterns that are not relevant to the diagnosis results, making the performance of the model weaker. Therefore, the noise removal of signal becomes a necessary step in the pre-processing stage. The purpose of ECG denoising is to eliminate three types of noises: electromyography interference, power-line interference and baseline wander. We used a notch filter to filter out the 60HZ power-line interference in the signal. Since the frequency of baseline wander noise ranges from 0.02hz to 2HZ, median filter was used to remove this kind of noise effectively. For high frequency noises such as electromyography, we used discrete wavelet transform to decompose the signals and filter out part of the detail components, then we used the principle of inverse wavelet transform to reconstruct the clean signal. As shown in Figure 1, the morphological characteristics of ECG signals can be well preserved.

Table 1. Details of the databases.

| Database                                      | Diagnosis | No. of records | fs     |
|-----------------------------------------------|-----------|----------------|--------|
| Long-Term ST Database                         | CHF       | 6              | 250HZ  |
| MGH/MF Waveform Database                      | CHF       | 2              | 360HZ  |
| Sudden Cardiac Death Holter Database          | CHF       | 2              | 250HZ  |
| Cerebral Vasoregulation in Elderly with Stroke Database | CHF       | 2              | 500HZ  |
| BIDMC Congestive Heart Failure Database       | CHF       | 15             | 250HZ  |
| PTB Diagnostic ECG Database                   | CHF       | 6              | 1000HZ |
| Fantasia Database                             | Normal    | 40             | 500HZ  |
| MIT-BIH Normal Sinus Rhythm Database          | Normal    | 18             | 128HZ  |

Figure 1. ECG signals after noises removal.

After completing the noise removal, we resampled the signal with sampling frequency greater or less than 250HZ to 250HZ, so as to unify the sampling frequency to the standard of 250HZ. Then, we performed R peak detection on the signals and extracted 300 points forward and 200 points backward to divide the ECG data as samples with the R peaks as the fiducial point. Due to the similarity of samples extracted in adjacent time, we randomly selected one sample at a certain interval (e.g., 20s) to ensure the diversity of the samples. Finally, we divided the database based on "inter-patient" pattern, i.e., 70% of patients were randomly selected from CHF and 70% of healthy people were randomly selected from Health Control to make up the training set, and the remaining 30% data constituted the testing dataset. The number of samples in training dataset and testing dataset are shown in table 2.

Table 2. The number of extracted samples.

| Diagnosis | Training set | Testing set |
|-----------|--------------|-------------|
| CHF       | 52287        | 26661       |
| Normal    | 77203        | 32934       |
2.3. Model structure

Convolutional neural network (CNN) has become a popular method in computer vision area. The milestone in the history of CNNs is the emergence of ResNet, which can train much deeper networks to achieve higher performance [10]. The idea of shot-cut connection in ResNet model has influenced many networks that appear latter. DenseNet draws on the idea of ResNet, and establishes connections between all the previous layers and later layers, hence its name [11]. Although the structure is not complicated, it is very effective. In this article, we introduced this idea into CHF detection by adapting the network structure to the diagnosis task, thereby improving the accuracy of the algorithm. The whole framework of our method is shown in Figure 2. The detailed information is given in Table 3 and the basic theories of each part of the structure are presented below.

![Figure 2. System Framework.](image)

2.3.1. First convolution layer. The input data is batch normalized through the Batch Normalization (BN) layer. Then it is convolved with filters with kernel size of 7 and stride of 2. In the first layer, the number of filters is $2k$, where $k$ is the growth rate. Next, a max pooling operation is performed with pooling size of 3 and stride size of 2 to sample the feature maps.

2.3.2. Dense Block layer. The structure of the "Dense Block" repeatedly appears in the network is shown in figure 3. The basic convolution operation in each Block adopts the same BN + Relu + Conv structure to ensure that the feature maps can be matched in the channel dimension. The feature maps are normalized by Batch Normalization layer, then activated by Relu activation function, and finally convolved with filters whose kernel size is 3 and number is $k$.

In each block, the input of one layer is calculated by contacting the output of all the former layers in channel dimension, and the output of this layer will also be passed to all the subsequent layers as input. The output of each layer is calculated by formula (1), where $[x_0, x_1, x_2, ..., x_{l-1}]$ is the contacted features from the former layers, and $l$ is the number of convolutional layers in each Block. Therefore,

$$x_l = H_l[x_0, x_1, x_2, ..., x_{l-1}]$$

(1)

the output channels are $k_0 + l \times k$ after a block and increase $l \times k$, where $k_0$ is input channels and $k$ is the growth rate.

![Figure 3. A Dense Block schematic architecture](image)
2.3.3. Transition layer. The BN + Conv + Avg pooling operations are used in the transition layer to match the size of feature maps between two blocks, where convolution makes the number of channels halved and average pooling was used to make the size of the feature maps halved.

2.3.4. Classification layer. After the last dense block, batch normalization, Relu activation and global average pooling operations are performed on the features extracted from the model to get the final higher-order features, and then output the final classification results using a fully-connected layer.

In this paper, we treated CHF patients as positive samples and healthy people as negative samples. The model classified the input data into these two categories. Therefore, binary cross-entropy was used as loss function, calculated by equation (2), where y was the true category of the sample and p was the predicted class of the network. \( \lambda \) was regularization rate to prevent the over-fitting problem.

\[
L(y, p) = -(y \log(p) + (1 - y) \log(1 - p)) + \lambda \sum ||w||^2
\]  

(2)

Table 3. The details of the model.

| Layers               | Output Size   | Operation                                              |
|----------------------|---------------|--------------------------------------------------------|
| Convolution          | 250 × 24      | 7 conv, stride=2                                        |
| Max Pooling          | 125 × 24      | 3 max pool, stride=2                                    |
| Dense Block (1)      | 125 × 96      | \[
|                     |               | [Batch Normalization] \times 6                          |
|                     |               | [Relu]                                                  |
|                     |               | [3 conv, stride=1]                                      |
| Transition Layer (1) | 125 × 48      | 1 conv, stride=1                                        |
|                     | 63 × 48       | 2 average pool, stride=2                                |
| Dense Block (2)      | 63 × 144      | \[
|                     |               | [Batch Normalization] \times 8                          |
|                     |               | [Relu]                                                  |
|                     |               | [3 conv, stride=1]                                      |
| Transition Layer (2) | 63 × 72       | 1 conv, stride=1                                        |
|                     | 32 × 72       | 2 average pool, stride=2                                |
| Dense Block (3)      | 32 × 216      | \[
|                     |               | [Batch Normalization] \times 12                         |
|                     |               | [Relu]                                                  |
|                     |               | [3 conv, stride=1]                                      |
| Transition Layer (3) | 16 × 108      | 1 conv, stride=1                                        |
|                     | 16 × 108      | 2 average pool, stride=2                                |
| Dense Block (4)      | 16 × 180      | \[
|                     |               | [Batch Normalization] \times 6                          |
|                     |               | [Relu]                                                  |
|                     |               | [3 conv, stride=1]                                      |
| Classification Layer | 1 × 108       | global average pooling                                  |
|                     |               | 1D fully-connected, Sigmoid activation                  |

2.4. Training parameter setting

In this study, the model was implemented in Tensorflow framework with Xavier initialization and zero initialization for weights and bias, respectively. After many comparison experiments, the optimal hyperparameters were determined to optimize the performance of the model. The model finally used SDG as optimizer with the learning rate set to 0.001, the batch size set to 64, the regularization rate set to 0.0001, and the maximum training steps set to 10000. The value of growth rate \( k \) in the network was set to 12, and the number of convolutional layers in each dense block were set to 6, 8, 12 and 6. Our model was trained on one graphics processing unit (GPU) NVIDIA Titan X with 12GB of graphics memory, and a total of approximately 2000s was needed to training the model.
3. Results

3.1. Performance evaluation method
Among all the ECG recognition algorithms, the performance was mainly evaluated by three indexes: accuracy, sensitivity and specificity. In order to compare with existing studies, we also used these three indexes to evaluate the strengths and weaknesses of the proposed model, which were calculated by equation (3), (4), (5), where TP represents the number of CHF diagnosed as CHF, TN represents the number of Normal correctly diagnosed as Normal, and FP, FN represent the number of Normal misclassified as CHF and CHF misclassified as Normal respectively.

\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)
\]

\[
\text{Se} = \frac{TP}{TP + FN} \quad (4)
\]

\[
\text{Sp} = \frac{TN}{TN + FP} \quad (5)
\]

In order to demonstrate the advantage of our method, we conducted multiple experiments to verify it. The experiments mainly included two aspects. First, due to most of current studies are based on the "intra-patient" mode to divide the data, the training dataset and testing dataset contain data from the same person, thus achieving high performance. Therefore, we divided the whole extracted samples into training set and testing set randomly with ratio of 3:1 and compared the experimental results with the previous researches. Second, considering that the idea of "training on old patients and testing on new patients" is more in line with the practical scenario, which was called "inter-patient" mode, we trained the model on the dataset containing 70% of patients and tested on the remaining data to give reliable evaluation indicators of the model.

3.2. Experimental results and discussion
We trained and tested the model according to the experimental schemes and parameter settings identified above. Table 4 lists the results of the experiments taken for comparison with existing studies. As we can see from table 4, our model can achieve higher performance without complexed feature extraction and feature selection works compared with the traditional machine learning algorithms based on feature extraction. At the same time, it can be seen that our proposed model comprehensively surpasses the existing deep learning-based CHF detection algorithms with an accuracy of 99.02%, which proves that our proposed methods has great superiority.

Table 4. Experimental results compared with the classic methods

| Author, Year | Methods | Metrics |
|--------------|---------|---------|
| Sharma et al. [3], 2018 | eigenvalue decomposition + SVM | Acc=93.33% Se=94.49% Sp=91.41% |
| Acharya et al. [4], 2017 | Empirical mode decomposition + SVM | Acc=97.64% Se=97.01% Sp=98.24% |
| Chen et al. [6], 2017 | SAE-based deep learning algorithm | Acc=72.86% Se=49.09% Sp=86.33% |
| Wang et al. [7], 2018 | LSTM neural network | Acc=85.13% Se=73.58% Sp=91.78% |
| Acharya et al. [8], 2018 | 11-layer CNN model | Acc=98.97% Se=99.01% Sp=98.87% |
| Our method, 2020 | Densely connected deep network | Acc=99.02% Se=99.11% Sp=98.90% |

Table 5 presents the experimental results based on the "inter-patient" data partition mode, where there is no intersection of patient data in training set and testing set, which is more in line with the practical application scenario and gives more reliable evaluation results. Figure 4 displays the relevant indicators change with the training steps, and it can be seen that our model converges fast, the loss decreases with the training process, and the accuracy increases with train steps. As can be seen in table 5, a larger proportion of CHF samples are misclassified as health, possibly because our dataset contains some data from patients with heart failure of grade I-II, whose symptoms are not obvious,
resulting in low sensitivity. Finally, the accuracy of our method is 94.97%, and the sensitivity and specificity are 89.38% and 99.50% respectively.

![Figure 4. Metrics over the training steps.](image)

**Table 5.** Confusion matrix for “inter-patient” based experiment result.

| Predicted | Health | CHF | Acc  | Se   | Sp   |
|-----------|--------|-----|------|------|------|
| Real      | Health | 32768 | 166  |      |      |
|           | CHF    | 2830  | 23831| 94.97% | 89.38% | 99.50% |

4. Conclusion
In this article, an effective algorithm was proposed to detect CHF using ECG signals. First, we built a CHF database on PhysioBank to better train the deep learning model. As far as I know, we are the first to build such a database which contains more diverse CHF patient data and can be used as a reference for future studies. Then, inspired by the idea of DenseNet, we applied this idea to CHF detection task. The experiment results demonstrated the performance of the proposed algorithm was superior to the existing studies. Finally, in order to evaluate the algorithm more objectively, we proposed an "inter-patient" pattern to divide the database. Our algorithm has good performance and the accuracy, sensitivity and specificity reaches 94.97%, 89.38% and 99.50%, respectively. In future work, a more effective network will be study to improve the accuracy of the CHF detection, especially the improvement of sensitivity. At the same time, the algorithm we studied will be deployed in mobile applications and medical products to realize real-time monitoring of CHF patients and provide reliable references for doctors’ decisions.

Acknowledgements
This work is supported by the National Natural Science Foundation of China (No. 61571268)

References
[1] Yancy C W et al. 2013 2013 ACCF/AHA Guideline for the Management of Heart Failure: Executive Summary A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines *Circulation* 128 pp 1810-52
[2] Ponikowski P et al. 2014 Heart failure: preventing disease and death worldwide *Ero. Soc. Cardiol* 1 pp 4-25
[3] Sharma R R, Kumar A, Pachori R B, Pachori R B and Acharya U R 2019 Accurate automated detection of congestive heart failure using eigenvalue decomposition based features extracted from HRV signals 39 pp 312-27
[4] Acharya U R et al. 2017 Application of empirical mode decomposition (EMD) for automated identification of congestive heart failure using heart rate signals *Neural Comput & Applic.* 28 pp 3073-94
[5] Rajpurkar P, Hannun A, Haghpanahi M, Bourn C and Ng A Y 2017 Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks

[6] Chen W, Liu G, Su S W, Jiang Q and Nguyen H T 2017 A CHF detection method based on deep learning with RR intervals. In: *international conference of the ieee engineering in medicine and biology society*, pp 3369-72

[7] Wang L, Zhou W, Liu N, Xing Y and Zhou X 2018 CHF Detection with LSTM Neural Network. In: *international conference of the ieee engineering in medicine and biology society*, pp 514-7

[8] Acharya U R, Fujita H, Oh S L, Hagiwara Y, Tan J H, Adam M and Tan R S J A I 2019 Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals *Applied Intelligence* 49 pp 16-27

[9] Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C and Stanley H E J C 2000 PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals *Circulation* 101 pp 215-20

[10] He K, Zhang X, Ren S and Sun J 2016 Deep Residual Learning for Image Recognition. In: *computer vision and pattern recognition*, pp 770-8

[11] Huang G, Liu Z, Der Maaten L V and Weinberger K Q 2017 Densely Connected Convolutional Networks. In: *computer vision and pattern recognition*, pp 2261-9