



Article

An Improved UNet++ Model for Congestive Heart Failure Diagnosis Using Short-Term RR Intervals

Meng Lei ¹ , Jia Li ¹, Ming Li ¹, Liang Zou ^{1,*}  and Han Yu ²

¹ School of Information and Electrical Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; lmsiee@cumt.edu.cn (M.L.); TS18060175P31@cumt.edu.cn (J.L.); liming@cumt.edu.cn (M.L.)

² School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU), Singapore 639798, Singapore; han.yu@ntu.edu.sg

* Correspondence: liangzou@cumt.edu.cn

Abstract: Congestive heart failure (CHF), a progressive and complex syndrome caused by ventricular dysfunction, is difficult to detect at an early stage. Heart rate variability (HRV) was proposed as a prognostic indicator for CHF. Inspired by the success of 2-D UNet++ in medical image segmentation, in this paper, we introduce an end-to-end encoder-decoder model to detect CHF using HRV signals. The developed model enhances the UNet++ model with Squeeze-and-Excitation (SE) residual blocks to extract deep features hierarchically and distinguish CHF patients from normal subjects. Two open-source databases are utilized for evaluating the proposed method, and three segment lengths of intervals between successive R-peaks are employed in comparison with state-of-the-art methods. The proposed method achieves an accuracy of 85.64%, 86.65% and 88.79% when 500, 1000 and 2000 RR intervals are utilized, respectively. It demonstrates that HRV evaluation based on deep learning can be an important tool for early detection of CHF, and may assist clinicians in achieving timely and accurate diagnoses.

Keywords: congestive heart failure; short-term RR intervals; UNet++



Citation: Lei, M.; Li, J.; Li, M.; Zou, L.; Yu, H. An Improved UNet++ Model for Congestive Heart Failure Diagnosis Using Short-Term RR Intervals. *Diagnostics* **2021**, *11*, 534. <https://doi.org/10.3390/diagnostics11030534>

Academic Editor: Leonid Chepelev

Received: 21 February 2021

Accepted: 14 March 2021

Published: 16 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Congestive heart failure (CHF) is the terminal stage of a variety of cardiovascular diseases, such as hypertension, coronary heart disease, and valvular heart disease [1]. Frequently-reported symptoms include left ventricular hypertrophy and left ventricular dilation, which may lead to neuroendocrine disorders and circulatory dysfunction [2]. If left untreated or not treated properly, heart failure will gradually worsen over time. According to epidemiological surveys around the world, for every 100 people, 3–5 people have different levels of heart failure. However, diagnosis is often difficult, especially for early diagnosis. Early diagnosis of CHF is able to slow down the progression prior to adverse events and improve the patients' chance of survival [3]. Moreover, the mortality rate in CHF patients within 5 years is as high as 50%, which makes it a major public health challenge [4]. Therefore, CHF diagnosis with high objectivity and reliability is highly desired.

Electrocardiogram (ECG), which contains abundant information of cardiac activities, is commonly utilized for heart rhythm analysis in hospitals [5]. However, manual ECG signal analysis is time-consuming, and the quality of diagnostic results depends on the clinician's expertise [6]. To address the limitations of manual ECG analysis, various automated CHF detection algorithms have been proposed. For instance, Gotsman et al. [7] found that QRS-T angle is relatively stable in patients with heart failure, and widening of the QRS-T angle has predictive value. The equal frequency in amplitude and equal width in time (EFiA-EWiT) discretization method was proposed to extract features from ECG signals, and a linear regression model was used to differentiate CHF from normal sinus rhythm

(NSR) patterns [6]. Acharya et al. developed a fully automatic CHF diagnosis model based on an 11-layer convolutional neural network which achieved an accuracy of 98.97% on a data set from PhysioBank [8].

RR interval, the time elapsed between successive R-peaks extracted from ECG signals, may contain useful information about heart diseases [9]. Previous studies show that heart rate variability (HRV), a phenomenon of the variation of heart rate in the time intervals, can serve as a biomarker for disease severity in patients with CHF [10]. Long-term HRV signals have been employed for CHF detection. Using Poincare plot, Tushar et al. [11] studied the differences of lag-response in CHF patients and normal subjects, and found that curvilinearity was lost in patients with CHF after exploring sequences up to 50,000 beats. A multistage classification method using non-equilibrium decision-tree-based support vector machine (DT-SVM) was proposed for risk assessment based on HRV [12]. Fifty four classical measures and 126 dynamic indices were extracted to detect and quantify CHF patients. Isler et al. [13] applied genetic algorithm (GA) to select the best features from the combination of standard HRV measures and wavelet entropy measures. They further employed the k-nearest neighbor classifier to distinguish 29 CHF patients from 54 healthy subjects.

Despite providing promising performance, existing methods always require long-term data which are difficult to acquire. For example, Yu et al. [14] proposed bispectral analysis and genetic algorithm for CHF detection based on long-term (24-h) HRV. Although their proposed method achieved a high accuracy of 96.38%, acquiring 24-h HRV data is not easy outside of the hospital environment, especially with mobile devices. Recently, various machine learning-based CHF detection methods have been developed with different lengths of short-term RR intervals and achieved inspiring performance. Liu et al. [15] proposed a multiscale entropy method to classify normal subjects from CHF patients and achieved accuracy of 85.5% and 85.6% when 1000 and 2000 RR intervals are used, respectively. Wang et al. [16] combined handcrafted features and deep-learning features, and utilized an ensemble classifier for CHF detection. Three types of RR segment length ($N = 500, 1000$ and 2000) were used to evaluate their proposed method with accuracy of 83.84%, 87.54%, and 85.71%, respectively.

Recently, deep learning algorithms are widely used in medical signal analysis and have achieved remarkable results [17,18]. Deep neural networks are able to automatically learn discriminative features from raw data. They have shown enormous potential in analyzing heterogeneous data and great generalization ability. For example, Chen proposed a sparse auto-encoder-based deep learning model in CHF detection using RR intervals [19]. Li et al. [20] combined deep neural network and distance distribution matrix to identify CHF and achieved an accuracy of 81.85%. Wang et al. [21] used long short-term memory (LSTM) deep network to detect CHF based on RR intervals and achieved 82.51%, 86.68% and 87.55% accuracy on $N = 500, 1000$ and 2000 RR intervals, respectively.

Although the above-mentioned studies have made significant progress, there are still two major limitations. Firstly, mainstream CHF detection algorithms are mostly based on handcrafted features. However, such approaches are not robust and rely too much on expert experience. To extract effective temporal features from HRV data, some researchers integrated time-domain, frequency-domain and nonlinear features together to form feature vectors [16], which is time-consuming and error-prone. In addition, such features generally are task-specific and lack generality. Meanwhile, errors in feature extraction process may be propagated to later stages, which negatively impact the detection performance. Although a few deep learning-based strategies were proposed for exploring the information of RR intervals, the performance is not satisfactory in comparison with approaches based on time-series ECG signals. Secondly, patients with mild CHF might be misclassified with healthy controls because the classification boundaries between them are not clear. Existing studies show that there is still great room for improvement, especially for mild CHF patients detection.

To address these problems, this paper presents an encoder-decoder model for CHF detection based on short-term RR intervals. The main contributions are summarized as follows:

1. This paper proposes a novel strategy to extract the features from short-term RR intervals, at most with the length of 2000. The end-to-end decision support system extracts deep features automatically from raw data, which not only improves the generalization but also avoids error propagation and information reduction.
2. Inspired by the success of UNet++ in computer vision tasks, this paper presents a novel strategy to classify patients with CHF and normal subjects via improved 1-D UNet++. To the best of our knowledge, no prior work on ECG analysis has employed the 1-D UNet++ network. In addition, to increase the sensitivity to informative features, the squeeze and excitation block are integrated [22] and combined with residual networks [23]. The improved 1-D UNet++ has achieved state-of-the-art accuracy of 88.79% on two publicly available data sets with mild CHF patients.

2. Materials and Methods

The proposed framework, consisting of data preprocessing, feature-extraction based on the improved UNet++ and classification based on fully connected layers, is illustrated in Figure 1. The details of each parts are discussed in the following sub-sections.

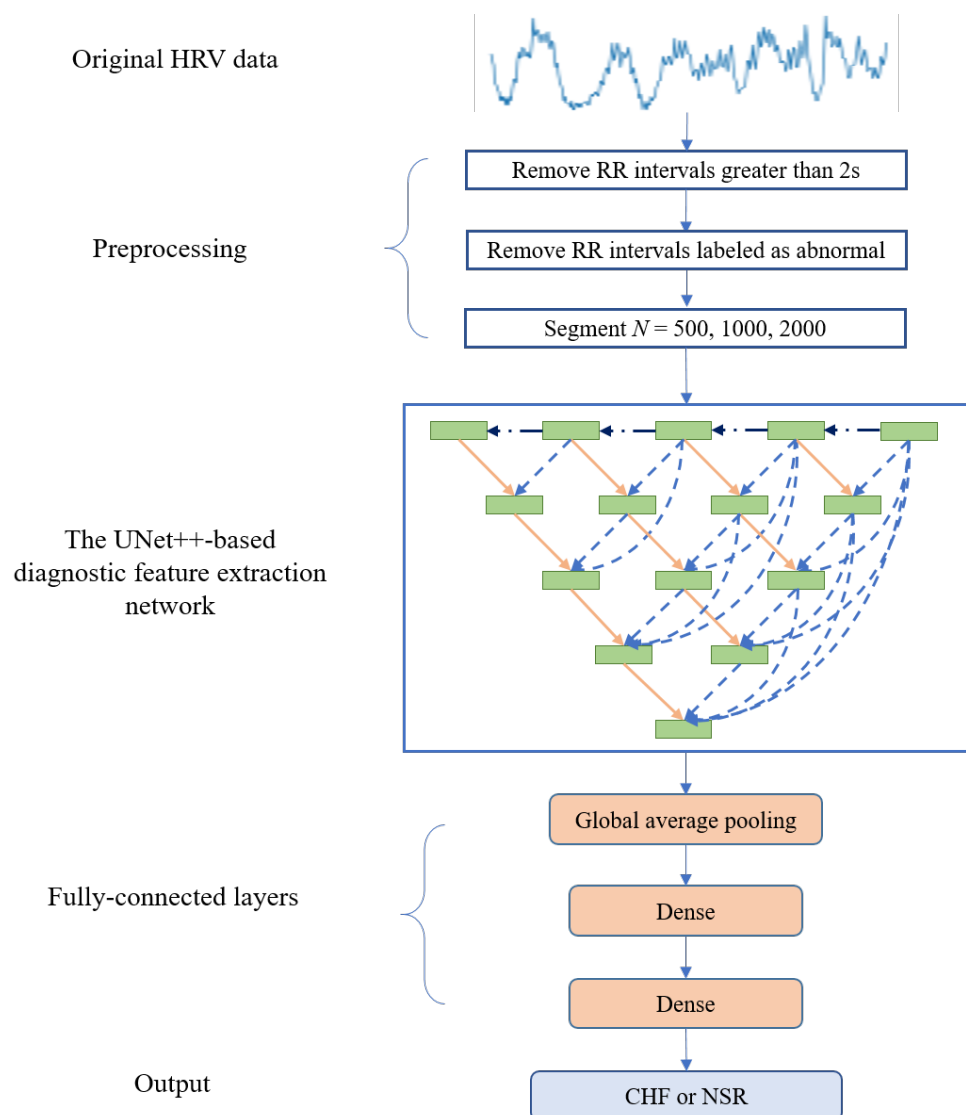


Figure 1. An overview of the method used in this work.

2.1. Dataset and Preprocessing

All data used in this study are obtained from PhysioBank [24], an open-access archive of physiological signals. For normal subjects, the normal sinus rhythm RR interval database (NSR-RR) is used. This database includes 54 long-term RR interval recordings from normal subjects aged from 29 to 76. The congestive heart failure RR interval database (CHF-RR) comprises 29 subjects with CHF (NYHA classes I, II and III) aged from 34 to 79. The raw ECG signals of both NSR-RR and CHF-RR databases were digitized at 128 samples per second, and the beat information was annotated through automated analysis with manual correction provided by PhysioNet [24]. RR interval is the time interval between successive cardiac cycle and has attracted wide attention for its potential to diagnose CHF. The pre-processing procedure for the RR intervals in this paper includes two steps:

1. Each beat in original ECG signals was annotated as normal (labeled as 'N') or abnormal (usually caused by the ectopic beats). The RR intervals marked as abnormal are removed to avoid the negative effects on analysis of HRV. Meanwhile, the RR intervals longer than 2 s are also removed to avoid the error accumulated in the precedent peak detection.
2. We split the ECG signal of each subject into multiple RR segments. This approach not only augments the data set, but also avoids the problem of time-consuming process in long-term HRV signal analysis [25]. To compare the results with other studies, the signals are divided into 500, 1000 and 2000 RR intervals, as in [15,16]. Table 1 shows the details for these two datasets used in this study. Two demonstrative examples of the signals corresponding to NSR and CHF with 500 RR intervals are shown in Figure 2.

Table 1. The number of signals for different database and classes.

Database	Pre-Processing	Total Segments		
		500	1000	2000
CHF-RR	None	6635	3317	1658
	Removing the RR intervals longer than 2 s	6622	3311	1655
	Removing the RR intervals marked as abnormal heartbeats	6271	3129	1558
NSR-RR	None	11,555	5777	2888
	Removing the RR intervals longer than 2 s	11,538	5769	2884
	Removing the RR intervals marked as abnormal heartbeats	11,314	5641	2808

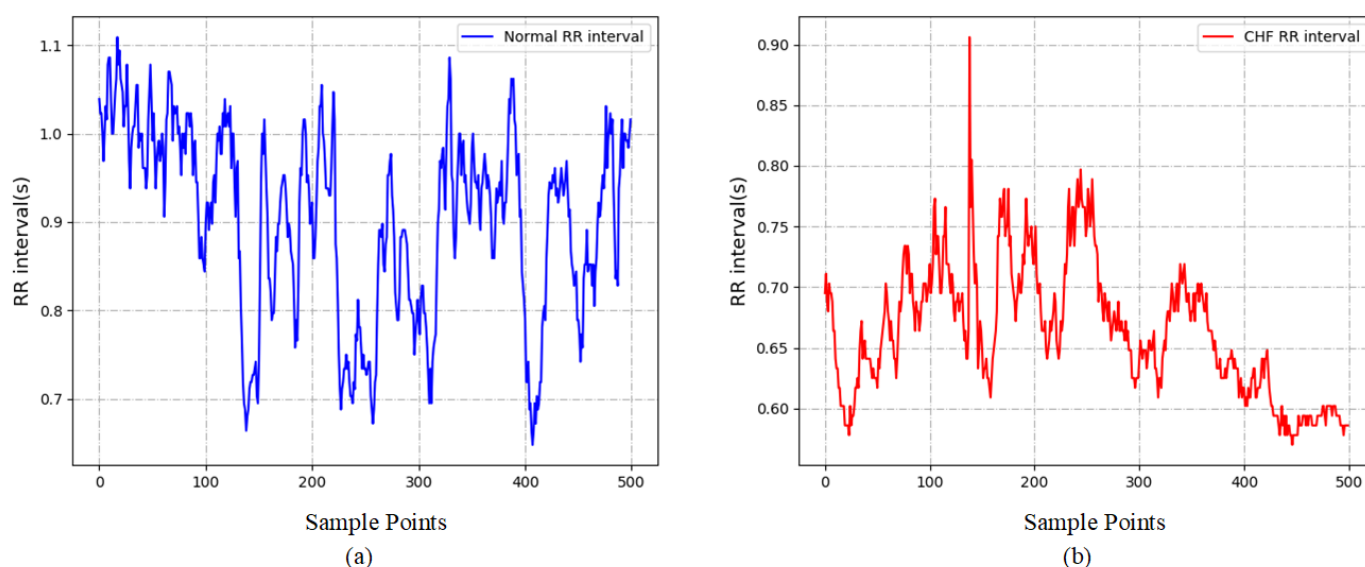


Figure 2. Two demonstrative examples for the RR intervals corresponding to (a) normal subjects and (b) CHF patients.

2.2. Proposed Network Architecture

The encoder-decoder architecture has become increasingly popular in feature extraction due to its high flexibility and superiority. UNet++ is a useful new variant of UNet, proposed by Zhou et al. [26,27]. A series of nested dense convolutional blocks connect the encoder and decoder in UNet++ , which can narrow and fill the information gap between the feature maps of the encoder and decoder prior to fusion. In this study, given that the CHF data is one-dimension, the 1-D variant of classical UNet++ model is developed to explore the pathological variations of CHF based on RR interval recordings. The 1-D UNet++ is able to capture valuable details of the HRV signals effectively since high dimension feature maps from the encoder part are gradually enriched prior to fusion with the corresponding pathologically rich feature maps from the decoder part.

RR intervals of different lengths, after being processed, are utilized as the input of the network. As shown in Figure 3a, the 1-D UNet++ structure consists of convolution blocks, down-sampling and up-sampling modules. Each black arrow denotes a down-sampling step which is implemented using a maximum pooling operation with kernel size of 2 and stride of 1. This window and stride configuration halves the size of the feature map. Such down-sampling can effectively extract features of the input and enhance robustness to noise by condensing features. The orange arrows represent the opposite operation, namely up-sampling, doubling the size of the feature maps. Up-sampling is the part that restore characteristics which is the highly effective expression form of the input data.

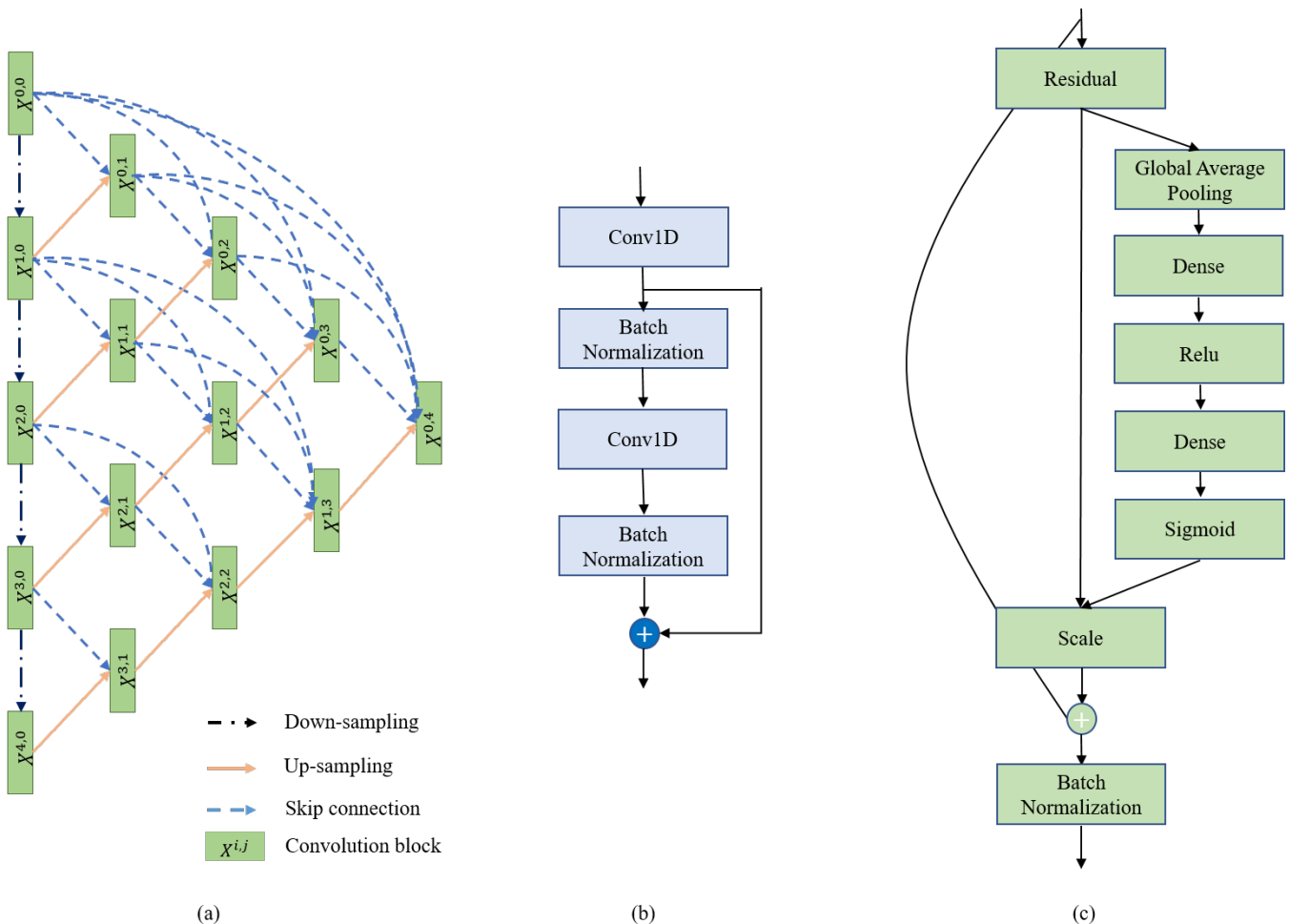


Figure 3. Structure of (a) the overall UNet++ network structure, (b) residual module and (c) convolution block.

As shown in Figure 3a, we assume $x^{i,j}$ denotes the output of node $X^{i,j}$, where i represents i th down-sampling layer along the encoder way and j represents the j th convolution layer along the skip pathway. The accumulation of feature maps by $x^{i,j}$ can be defined as:

$$x^{i,j} = \begin{cases} H(x^{i-1,j}), & j = 0 \\ H([\![x^{(i,k)}]_{k=0}^{j-1}\!] u(x^{i+1,j-1})), & j > 0 \end{cases} \quad (1)$$

where $H(\cdot)$ denotes a 1-D convolution operation combined with an activation function, $u(\cdot)$ presents an up-sampling layer, and $[\cdot]$ is the concatenation operation. Generally, nodes at level $j = 0$ receive only one input from the previous layer of the encoder while nodes at level $j > 0$ receive $j + 1$ inputs from both the skip connections and the up-sampling layer.

Residual modules are incorporated in the convolution unit, which facilitates convergence of the proposed deep model [28]. As can be seen in Figure 3b, the original residual module contains 1-D convolution layers (Conv1D) [29] and Batch Normalization (BN) layers [30] which are implemented alternately. The output of the residual module is generated by adding the outputs of the first Conv1D layer and the second BN layer. Inspired by the effectiveness of squeeze-and-excitation features on image object classification, 1-D squeeze-and-excitation (SE) residual modules are employed as convolution units in UNet++, as shown in Figure 3c. The SE residual modules can adaptively recalibrate residual feature maps within each feature channel by explicitly modeling interdependency between channels. It can enhance the representational power of modules throughout the network [22].

2.3. Model Structure and Parameters

In this work, a deep learning model is built to perform CHF detection using RR interval recordings. As shown in Figure 3a, the dimension of output to encoding is 1/16 of the input size. In this work, normal and CHF recordings, each with 500, 1000 or 2000 samples, are changed to 512, 1008 and 2000 using zero padding, respectively and then are fed into the input layer of this model. A global average pooling [31] is used to summarize the information from all the feature maps. Finally, an automatic prediction is provided by learning of these feature maps in the dense layer.

The signals from 83 subjects are split into 10 parts and ensure that all the signals from each subject are in one part. 10-fold cross validation is employed to evaluate the robustness of the proposed model. For each iteration, 9 parts are used for training and the remaining one part is used for testing. The method is repeated 10 times by shifting the testing part. The test set consists of the RR intervals from the subjects who are not used in the training process, and therefore reduces the possibility of over-fitting. We empirically set the batch size as 16 and the number of training epochs as 70 ($N = 500, 1000$ and 2000 length RR intervals). During the training process, we first set the initial learning rate as 10^{-4} , and update its value as 0.1 times of the original one when the validation loss stops improving within 5 epochs. Adam optimizer and mean squared error loss function are applied. The loss function is defined as:

$$L(y, p) = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2 \quad (2)$$

where N is the number of samples, y_i and p_i is the true label and the prediction result of the i th sample.

2.4. Performance Measures

To evaluate the performance of the proposed method and make a fair comparison, we employ four widely used evaluation metrics, including accuracy, recall, precision and F1-score [32,33]. These four indicators are widely considered to be the most informative for evaluating the performance of classifiers and convenient for calculation. All these evaluation metrics can be calculated by following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$F1 - score = \frac{2}{\left[\frac{1}{Precision} + \frac{1}{Recall}\right]} \quad (6)$$

Here, true positives (TP) is the number of CHF segments correctly classified as CHF group; false positives (FP) is the number of NSR segments wrongly classified as CHF group. True negatives (TN) associates with the number of NSR segments correctly classified as NSR group; false negatives (FN) is the number of CHF segments wrongly classified as NSR group.

3. Results

3.1. 10-Fold Cross Validation Performance

In the traditional cross validation, the RR intervals from one subject may appear in both training and test set. However, the similarity between these signals will lead to information leakage and result in overoptimistic performance. Considering the practical application of classification system to diagnosis unknown subjects, the dataset is split into training and test set in terms of the subjects in this study. The RR intervals of one subject appear in either the training set or the test set in each iteration.

The training details in terms of the accuracy and the loss against each epoch are presented in Figure 4. The solid line is the average of the performance across 10-folds. In the training phase, the model reaches convergence within a short time (mostly 20 epochs). The fluctuations of the performance on the test set is relative small, which demonstrates that the developed model generalizes well on separate dataset. Specifically, the training accuracy converges at 92.82%, 93.36% and 93.79% when 500, 1000 and 2000 RRs are employed respectively. The corresponding test accuracy converges at 85.64%, 86.65% and 88.79%.

3.2. Comparison with Different Network Architectures

Inception was a deep learning network model, designed by Christian Szegedy and others [34]. It is not only able to efficiently reduce the number of parameters, but also capable of increasing the expression ability of the network by introducing more linear mappings. In this study, the CHF detection results of the improved UNet++ are compared with those of introducing the inception modules on account of the excellent performance of Inception. In addition, the performance of the UNet++ model with or without SE modules are evaluated for illustrating its effect.

The results and comparisons of three RR segment length types ($N = 500, 1000$ and 2000) for different methods are listed in Table 2. The testing accuracy of the proposed UNet++ model among all the RR intervals length is over 85%. The best classification performance is yielded when 2000 RR intervals from each subject are employed, reaching an accuracy of 88.79%. These outperforming results indicate that the proposed model is effective for CHF detection.

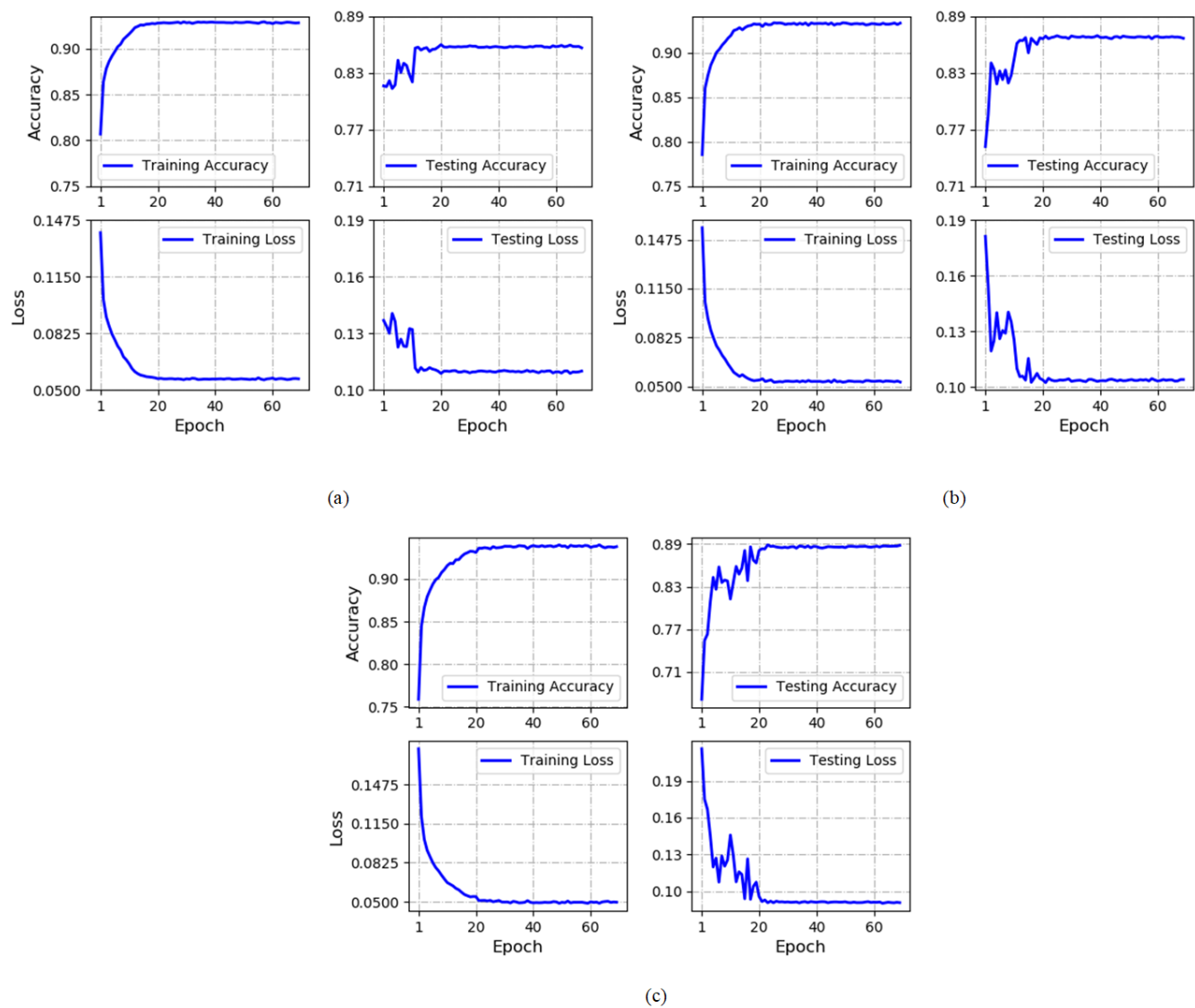


Figure 4. Performance graphs of the proposed model. (a) The RR segment length $N = 500$; (b) The RR segment length $N = 1000$; (c) The RR segment length $N = 2000$.

Table 2. The overall performance with different length of RR segments.

Methods	Segment Length	Evaluation			
		Recall	Precision	F1-Score	Accuracy
The 1-D UNet++	500	0.6812	0.7927	0.7248	0.8257
	1000	0.6621	0.7980	0.7176	0.8184
	2000	0.6617	0.8202	0.7279	0.8269
The inception 1-D UNet++	500	0.6978	0.8255	0.7488	0.8412
	1000	0.7247	0.8359	0.7684	0.8521
	2000	0.7190	0.8122	0.7529	0.8442
The proposed method	500	0.7381	0.8346	0.7793	0.8564
	1000	0.7596	0.8488	0.7947	0.8665
	2000	0.8018	0.8685	0.8281	0.8879

It is worth noting that the improved UNet++ with SE residual units consistently outperforms other network structures when three different lengths of RR intervals are employed, in terms of Recall, Precision, F1-score and Accuracy. The mean accuracy of the plain 1-D UNet++ across 10-fold across validation is 82.57%, 81.84% and 82.69% for $N = 500$, 1000 and 2000, respectively, whereas that of improved 1-D UNet++ is 85.64%, 86.65% and 88.79%, respectively. By introducing SE residual blocks to the original 1-D UNet++, the mean accuracy is increased by 3.07% 4.81% and 6.10% when $N = 500$, 1000 and 2000, respectively.

Except for the four aforementioned performance metrics, Receiver Operating Characteristic (ROC) curve and Area Under ROC Curve (AUC) [32] are also employed to evaluate the CHF detection performance. The ROC curve is able to accurately reflect the relationship between true positive rate (TPR) and the false positive rate (FPR) in a graphical way and is a comprehensive representative of the detection accuracy. AUC is obtained by summing the areas of the parts under the ROC curve. As can be seen from Figure 5, the UNet++ with SE residual modules, which is blue line, achieve the highest AUC values. The corresponding AUC is 0.90, 0.91 and 0.92 on $N = 500$, 1000 and 2000 respectively. Such experimental results demonstrate that the proposed model yields outstanding performance. In addition, the ROC curves of the UNet++ model without SE modules and inception UNet++ are shown in red and brown, respectively.

As shown in Figure 5, the AUC values of 1-D UNet++ without SE modules (AUC = 0.86, 0.86, and 0.88 when $N = 500$, 1000 and 2000, respectively) are less than those of the UNet++ with SE residual modules (AUC = 0.90, 0.91, and 0.92 when $N = 500$, 1000 and 2000, respectively). It mainly owns to that the UNet++ model with SE residual modules can efficiently exploit the information from RR interval segments. SE modules is able to learn the channel-wise calibration and help alleviate the dependencies among channel-wise features. Besides, the skip connection in SE residual blocks is conducive to the back-propagation of gradients and mitigates the declining-accuracy phenomenon in the deep network [35]. The inception modules can improve the expression ability of network by organizing information across channels. Both Table 2 and Figure 5 show that the UNet++ with inception modules do improve the performance in comparing with the plain UNet++ model whereas it is less effective than the improved UNet++ model with SE residual modules.

3.3. Comparison with State-of-the-Art CHF Diagnosis Methods

Over the past years, there are a variety of automatic classifiers to diagnose patients with CHF (Table 3). In this experiment, the proposed method is compared with several state-of-the-art CHF diagnosis methods, including the methods based on Inception-V4 [20], LSSVM [36], SVM [15], Ensemble classifier [16] and LSTM [21] in terms of the diagnosis accuracy. Li et al. [20] obtained 81.85% with 300 length RR interval segments. Sharma et al. [36] achieved the accuracy of 87.15% by using $N = 2000$ length RR intervals for the classification of normal and CHF signals. Liu et al. [15] studied CHF detection and obtained the performance of 85.5% and 85.6% on $N = 1000$ and 2000 length RR intervals, respectively. Wang et al. [16] combined expert features of RR intervals with deep-learning features, and fed into the ensemble classifier to differentiate CHF patients from healthy controls. They yielded 83.84%, 87.54% and 85.71% accuracy on $N = 500$, 1000 and 2000 length RR intervals, respectively. Wang et al. [21] presented an LSTM-based inception module to detect CHF and achieved 82.51%, 86.68% and 87.55% accuracy on $N = 500$, 1000 and 2000 length RR intervals, respectively.

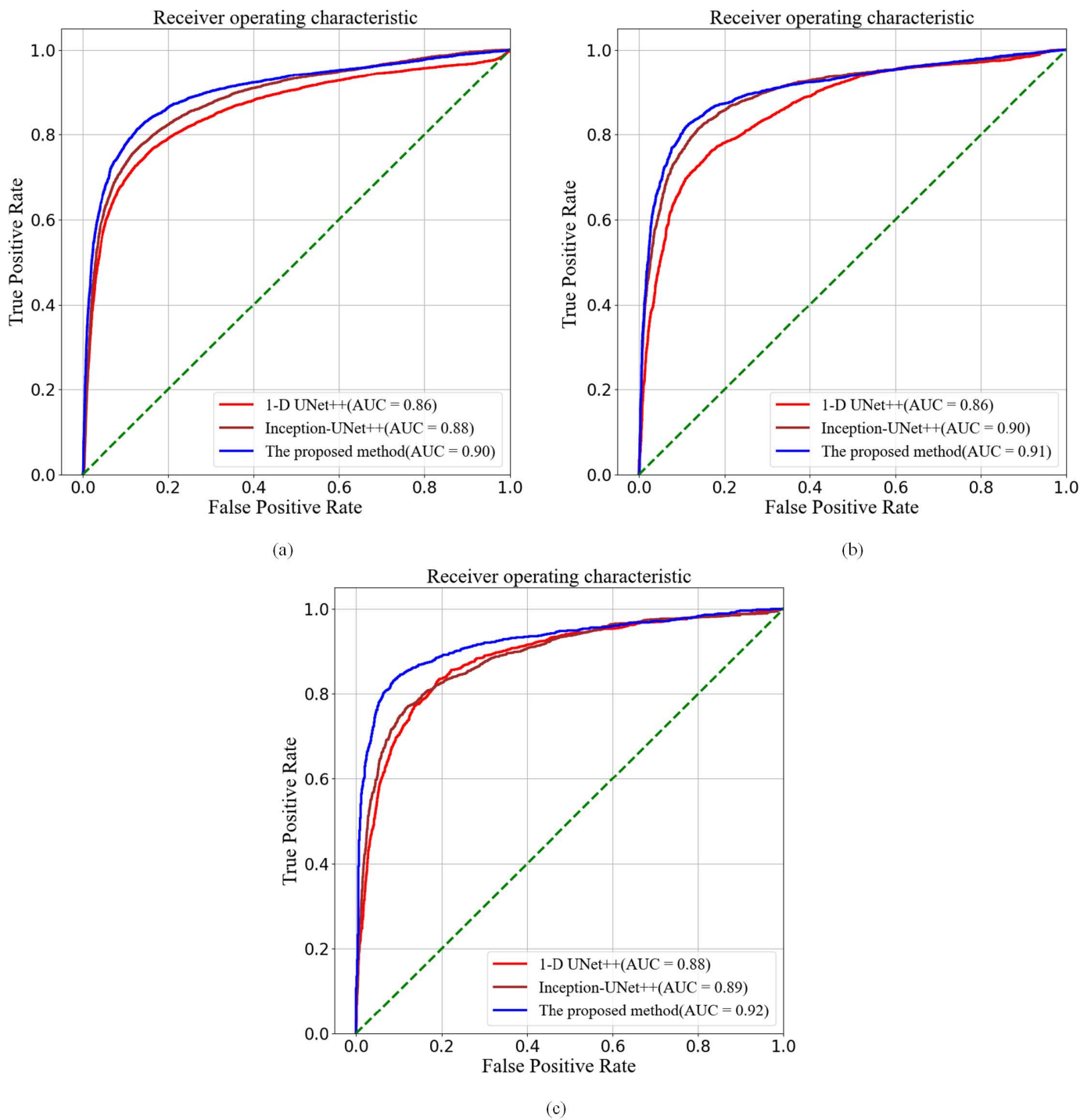


Figure 5. The ROC curve of CHF detection. (a) The RR segment length $N = 500$; (b) The RR segment length $N = 1000$; (c) The RR segment length $N = 2000$.

In [21], Wang et al. employed the traditional 10-fold cross validation and achieved the mean accuracy of 86.42%, 87.76% and 86.63% with $N = 500$, 1000 and 2000 length RR intervals, respectively. However, the shuffled signals of all the patients were divided them into training set and test set. Such division method is against the inter-patient experiment of Association for the Advancement of Medical Instrumentation (AAMI) standard [37]. To overcome such issue, they also employed the blindfold testing to evaluate the result [21]. They randomly selected the RR intervals of 12 subjects as the test data, and achieved an accuracy of 82.51%, 86.68% and 87.55% when $N = 500$, 1000 and 2000 respectively. In this study, to demonstrate how well the model perform on unseen data, we show the ROC

curves corresponding to 10 folds when 2000 intervals are employed, as in Figure 6. It is clear that the AUCs vary greatly from fold to fold. The largest AUC is 1 with accuracy of 95.95% (fold 2) whereas the smallest AUC is only 0.75 with accuracy of 78.54% (fold 3).

Table 3. Comparison of the proposed method against existing methods on CHF detection.

Author (year)	Classifier	Features	Length	Accuracy
Li (2018) [20]	Inception-V4	Fuzzy GMEn	300	81.85%
Sharma (2018) [36]	LS-SVM	k-NN entropy and correntropy	2000	87.15%
Liu (2017) [15]	SVM	Multiscale entropy of RR	1000	85.5%
			2000	85.6%
Wang (2019) [16]	Ensemble classifier	Expert features and deep-learning features	500	83.84%
			1000	87.54%
			2000	85.71%
Wang (2019) [21]	LSTM based Inception	-	500	82.51%
			1000	86.68%
			2000	87.55%
Our proposed method	Improved UNet++	-	500	85.64%
			1000	86.65%
			2000	88.79%

Compared with other methods, the proposed model achieves the best performance of 85.64%, 86.65% and 88.79% when $N = 500$, 1000 and 2000, respectively. One potential reason is that the proposed method is able to extract more reliable signal features in high dimensional space. UNet ++ shortens the information gap between encoder and decoder through the information fusion between different layers, which makes full use of RR signals. Furthermore, the SE residual blocks is able to emphasize the salient features and suppress the irrelevant information [22].

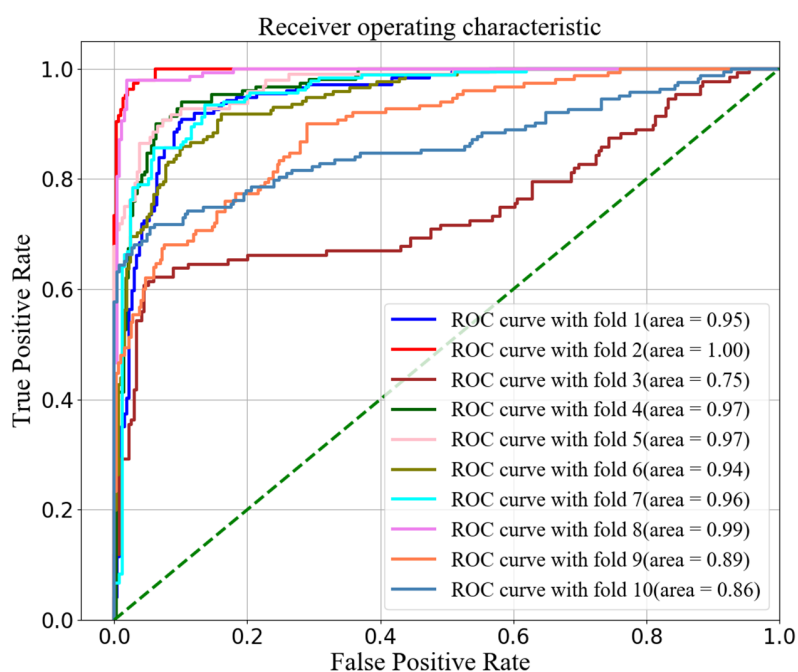


Figure 6. The ROC curve of 10 folds with 2000 sample length.

3.4. Performance Evaluation in More Practical Scenario

In real-world applications, clinicians need to differentiate CHF subjects with non-CHF subjects, rather than only the normal subjects. Therefore, to fairly demonstrate the performance of the proposed method in realistic scenario, other types of heart rhythm abnormalities should be considered. In this study, we further evaluate the performance when the RR intervals of atrial fibrillation (AF) patients are also employed as the non-CHF signals. These ECG signals are from public-available long-term AF database which includes 84 long-term (24-h) ECG recordings [38]. After pre-processing, 4333 RR segments of AF (when the segment length $N = 2000$) are obtained. To avoid the problem of class imbalance, we mixed the signals from the Normal Sinus Rhythm (NSR) RR interval Database and long-term AF database, and then randomly selected 2800 RR segments from them. Table 4 illustrates the details of these datasets.

Table 4. The number of ECG recordings in each dataset when the segment length $N = 2000$.

Classes	Database	Total Segments ($N = 2000$)
Non-CHF	NSR RR interval Database	2808
	Long-Term AF Database	4333
CHF	After random sampling	2800
	CHF RR interval Database	1558

As shown in Table 5, the average accuracy after 70 epochs is 89.33% when the RR intervals of both NSR and AF subjects are utilized as the non-CHF data. The result is similar to that in differentiating NSR and CHF patients. However, this study is a preliminary attempt to automatically diagnose CHF, and many other types of heart rhythm abnormalities will be considered in the future research.

Table 5. The overall performance with different kinds of non-CHF signals when the segment length $N = 2000$.

Data	Evaluation			
	Recall	Precision	F1-Score	Accuracy
CHF vs. NSR	80.18%	86.85%	82.81%	88.79%
CHF vs. NSR and AF	78.67%	88.26%	82.24%	89.33%

4. Discussion and Conclusions

In this work, an automatic classifier for CHF diagnosis via short length HRV signals is proposed. In comparison with previous CHF detection methods, the developed method employs an end-to-end deep learning model to extract features and make decision. To be more specific, the improved 1-D UNet++ architecture involves a residual block to distinguish CHF patients from normal subjects as well as a SE block to highlight the useful features and suppress the useless information. Such classification model obtains information from HRV signals with minimal information reduction and provides the optimal feature of the input RR interval segments. The proposed model outperformed the previous CHF diagnosis with a state-of-art accuracy of 85.64%, 86.65% and 88.79% when 500, 1000 and 2000 RR intervals are employed, respectively. This pilot study demonstrates that the deep learning-based automatic diagnosis can be an important tool to assist clinicians in making wise decisions. Moreover, CHF diagnosis via short-term RR intervals can be transplanted to mobile devices like smartphones easily. It contributes to monitoring the changes of cardiac autonomic nervous function with CHF patients.

Although the proposed method has provided promising results, there are still a few limitations to overcome. First, more training data, especially from many other types of heart rhythm abnormalities, is required to provide more reliable diagnosis. Second, CHF

is categorized into four stages by the American College of Cardiology Foundation [39]. The proposed method can only determine whether the patients suffer from CHF or not, but cannot determine the precise CHF stage due to limited numbers of subjects available in stage I and II CHF.

Author Contributions: Conceptualization, M.L. (Meng Lei) and J.L.; Funding acquisition, L.Z.; Methodology, J.L. and M.L. (Ming Li); Project administration, H.Y.; Supervision, M.L. (Ming Li) and L.Z.; Validation, L.Z.; Writing—original draft, M.L. (Meng Lei) and J.L.; Writing—review & editing, L.Z., and H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities with grant number 2019ZDPY17.

Institutional Review Board Statement: Ethical review and approval were waived for this study, because this study utilizes public data.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Cai, A.; Zhang, J.; Wang, R.; Chen, J.; Huang, B.; Zhou, Y.; Wang, L. Joint effects of obstructive sleep apnea and resistant hypertension on chronic heart failure: A cross-sectional study. *Int. J. Cardiol.* **2018**, *257*, 125–130. [[CrossRef](#)]
- Baba, H.A.; Wohlschlaeger, J. Morphological and molecular changes of the myocardium after left ventricular mechanical support. *Curr. Cardiol. Rev.* **2008**, *4*, 157. [[CrossRef](#)]
- Zamfirescu, M.B.; Ghilencea, L.N.; Popescu, M.R.; Bejan, G.C.; Ghiordanescu, I.M.; Popescu, A.C.; Dorobanțu, S.G. A Practical Risk Score for Prediction of Early Readmission after a First Episode of Acute Heart Failure with Preserved Ejection Fraction. *Diagnostics* **2021**, *11*, 198. [[CrossRef](#)] [[PubMed](#)]
- Wu, M.Y.; Chen, T.T.; Wu, M.S.; Tu, Y.K. Radio-contrast medium exposure and dialysis risk in patients with chronic kidney disease and congestive heart failure: A case-only study. *Int. J. Cardiol.* **2021**, *324*, 199–204. [[CrossRef](#)] [[PubMed](#)]
- Martín-Yebra, A.; Monasterio, V.; Cygankiewicz, I.; Bayés-de Luna, A.; Caiani, E.G.; Laguna, P.; Martínez, J.P. Post-ventricular premature contraction phase correction improves the predictive value of average T-wave alternans in ambulatory ECG recordings. *IEEE Trans. Biomed. Eng.* **2017**, *65*, 635–644.
- Orhan, U. Real-time CHF detection from ECG signals using a novel discretization method. *Comput. Biol. Med.* **2013**, *43*, 1556–1562. [[CrossRef](#)] [[PubMed](#)]
- Gotsman, I.; Shauer, A.; Elizur, Y.; Zwas, D.R.; Lotan, C.; Keren, A. Temporal changes in electrocardiographic frontal QRS-T angle and survival in patients with heart failure. *PLoS ONE* **2018**, *13*, e0194520. [[CrossRef](#)] [[PubMed](#)]
- Acharya, U.R.; Fujita, H.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M.; San Tan, R. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Appl. Intell.* **2019**, *49*, 16–27. [[CrossRef](#)]
- Andršová, I.; Hnatkova, K.; Šišáková, M.; Toman, O.; Smetana, P.; Huster, K.M.; Barthel, P.; Novotný, T.; Schmidt, G.; Malik, M. Heart Rate Influence on the QT Variability Risk Factors. *Diagnostics* **2020**, *10*, 1096. [[CrossRef](#)]
- Sharma, V.; Lonie, S.; Ta, D.; Toia, D.; Selig, S.; Hare, D. Inter-tester Reliability of Heart Rate Variability in CHF Patients. *Heart Lung Circ.* **2012**, *21*, S254. [[CrossRef](#)]
- Thakre, T.P.; Smith, M.L. Loss of lag-response curvilinearity of indices of heart rate variability in congestive heart failure. *BMC Cardiovasc. Disord.* **2006**, *6*, 27. [[CrossRef](#)]
- Chen, W.; Zheng, L.; Li, K.; Wang, Q.; Liu, G.; Jiang, Q. A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement. *PLoS ONE* **2016**, *11*, e0165304. [[CrossRef](#)] [[PubMed](#)]
- İşler, Y.; Kuntalp, M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Comput. Biol. Med.* **2007**, *37*, 1502–1510. [[CrossRef](#)]
- Yu, S.N.; Lee, M.Y. Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. *Comput. Biol. Med.* **2012**, *42*, 816–825. [[CrossRef](#)]
- Liu, C.; Gao, R. Multiscale entropy analysis of the differential RR interval time series signal and its application in detecting congestive heart failure. *Entropy* **2017**, *19*, 251.
- Wang, L.; Zhou, W.; Chang, Q.; Chen, J.; Zhou, X. Deep ensemble detection of congestive heart failure using short-term rr intervals. *IEEE Access* **2019**, *7*, 69559–69574. [[CrossRef](#)]
- Zou, L.; Zheng, J.; Miao, C.; Mckeown, M.J.; Wang, Z.J. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. *IEEE Access* **2017**, *5*, 23626–23636. [[CrossRef](#)]
- Cui, H.; Liu, A.; Zhang, X.; Chen, X.; Wang, K.; Chen, X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowl.-Based Syst.* **2020**, *205*, 106243. [[CrossRef](#)]

19. Chen, W.; Liu, G.; Su, S.; Jiang, Q.; Nguyen, H. A CHF detection method based on deep learning with RR intervals. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 3369–3372.
20. Li, Y.; Zhang, Y.; Zhao, L.; Zhang, Y.; Liu, C.; Zhang, L.; Zhang, L.; Li, Z.; Wang, B.; Ng, E.; et al. Combining convolutional neural network and distance distribution matrix for identification of congestive heart failure. *IEEE Access* **2018**, *6*, 39734–39744. [[CrossRef](#)]
21. Wang, L.; Zhou, X. Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. *Sensors* **2019**, *19*, 1502. [[CrossRef](#)]
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
23. Zhang, K.; Sun, M.; Han, T.X.; Yuan, X.; Guo, L.; Liu, T. Residual networks of residual networks: Multilevel residual networks. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 1303–1314. [[CrossRef](#)]
24. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)]
25. Smith, A.L.; Owen, H.; Reynolds, K.J. Heart rate variability indices for very short-term (30 beat) analysis. Part 1: Survey and toolbox. *J. Clin. Monit. Comput.* **2013**, *27*, 569–576. [[CrossRef](#)]
26. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [[CrossRef](#)]
27. Lei, M.; Rao, Z.; Wang, H.; Chen, Y.; Zou, L.; Yu, H. Maceral groups analysis of coal based on semantic segmentation of photomicrographs via the improved U-net. *Fuel* **2021**, *294*, 120475. [[CrossRef](#)]
28. Peng, D.; Zhang, Y.; Guan, H. End-to-end change detection for high resolution satellite images using improved unet++. *Remote Sens.* **2019**, *11*, 1382. [[CrossRef](#)]
29. Goh, C.H.; Tan, L.K.; Lovell, N.H.; Ng, S.C.; Tan, M.P.; Lim, E. Robust PPG Motion Artifact Detection Using a 1-D Convolution Neural Network. *Comput. Methods Programs Biomed.* **2020**, *196*, 105596. [[CrossRef](#)] [[PubMed](#)]
30. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 7–9 July 2015; pp. 448–456.
31. Li, Z.; Wang, S.H.; Fan, R.R.; Cao, G.; Zhang, Y.D.; Guo, T. Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. *Int. J. Imaging Syst. Technol.* **2019**, *29*, 577–583. [[CrossRef](#)]
32. Yu, B.; Zhang, X.; Wu, L.; Chen, X.; Chen, X. A novel postprocessing method for robust myoelectric pattern-recognition control through movement pattern transition detection. *IEEE Trans. Hum.-Mach. Syst.* **2019**, *50*, 32–41. [[CrossRef](#)]
33. Zou, L.; Yu, X.; Li, M.; Lei, M.; Yu, H. Nondestructive identification of coal and gangue via near-infrared spectroscopy based on improved broad learning. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8043–8052. [[CrossRef](#)]
34. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
35. Zhu, W.; Huang, Y.; Zeng, L.; Chen, X.; Liu, Y.; Qian, Z.; Du, N.; Fan, W.; Xie, X. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* **2019**, *46*, 576–589. [[CrossRef](#)] [[PubMed](#)]
36. Sharma, R.R.; Kumar, A.; Pachori, R.B.; Acharya, U.R. Accurate automated detection of congestive heart failure using eigenvalue decomposition based features extracted from HRV signals. *Biocybern. Biomed. Eng.* **2019**, *39*, 312–327. [[CrossRef](#)]
37. Stergiou, G.S.; Alpert, B.; Mieke, S.; Asmar, R.; Atkins, N.; Eckert, S.; Frick, G.; Friedman, B.; Grafl, T.; Ichikawa, T.; et al. A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement. *Hypertension* **2018**, *71*, 368–374. [[CrossRef](#)] [[PubMed](#)]
38. Petrutiu, S.; Sahakian, A.V.; Swiryn, S. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace* **2007**, *9*, 466–470. [[CrossRef](#)] [[PubMed](#)]
39. Koglin, J.; Pehlivanli, S.; Schwaiblmair, M.; Vogeser, M.; Cremer, P.; von Scheidt, W. Role of brain natriuretic peptide in risk stratification of patients with congestive heart failure. *J. Am. Coll. Cardiol.* **2001**, *38*, 1934–1941. [[CrossRef](#)]