

Received March 13, 2019, accepted April 12, 2019, date of publication April 22, 2019, date of current version June 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912226

Deep Ensemble Detection of Congestive Heart Failure Using Short-Term RR Intervals

LUDI WANG¹, WEI ZHOU², QING CHANG³, JIANGEN CHEN⁴, AND XIAOGUANG ZHOU^{1,5}

¹Automation School, Beijing University of Posts and Telecommunications, Beijing 100876, China

²Department of neuroscience, Uppsala University, 751 24 Uppsala, Sweden

³Shanghai General Practice Medical Education and Research Center, Jiading District Central Hospital Affiliated Shanghai University of Medicine and Health Sciences, Shanghai 201800, China

⁴Department of General Medicine, Jiading Industrial District Community Health Service Center, Shanghai 201800, China

⁵School of Economics and Management, Minjiang University, Fuzhou 350108, China

Corresponding author: Xiaoguang Zhou (zgx@bupt.edu.cn)

This work was supported in part by the Open Foundation of State key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, under Grant SKLNST-2018-1-18, and in part by the BUPT Excellent Ph.D. Students Foundation.

ABSTRACT Heart rate variability (HRV) is an effective predictor of congestive heart failure (CHF). However, important challenges exist regarding the effective temporal feature extraction and efficient classification using high-dimensional HRV representations. To solve these challenges, an ensemble method for CHF detection using short-term HRV data and deep neural networks was proposed. In this paper, five open-source databases, the BIDMC CHF database (BIDMC-CHF), CHF RR interval database (CHF-RR), MIT-BIH normal sinus rhythm (NSR) database, fantasia database (FD), and NSR RR interval database (NSR-RR), were used. Additionally, three RR segment length types ($N = 500, 1000$, and 2000) were used to evaluate the proposed method. First, we extracted the expert features of RR intervals (RRIs) and then built a long short-term memory-convolutional neural network-based network to extract deep-learning (DL) features automatically. Finally, an ensemble classifier was used for CHF detection using the above features. With blindfold validation (three CHF subjects and three normal subjects), the proposed method achieved 99.85%, 99.41%, and 99.17% accuracy on $N = 500, 1000$, and 2000 length RRIs, respectively, using the BIDMC-CHF, NSR, and FD databases. With blindfold validation (six CHF subjects and six normal subjects), the proposed method achieved 83.84%, 87.54%, and 85.71% accuracy on $N = 500, 1000$, and 2000 length RRIs, respectively, using the NSR-RR and CHF-RR databases. Based on feature ranking, the significant effectiveness provided by the DL features has been proven. The results have shown that the deep ensemble method can achieve reliable CHF detection using short-term heart rate signals and enable CHF detection through intelligent hardware.

INDEX TERMS Electrocardiography, boosting, artificial intelligence.

I. INTRODUCTION

Heart failure (HF) is a clinical syndrome of various heart diseases at severe stages and is also known as congestive HF (CHF). This condition is caused by the inadequate blood filling function of the ventricular pump. A poor heart pump function can cause an insufficient heart discharge volume that fails to meet the needs of body metabolism. Additionally, the blood perfusion of tissues and organs becomes insufficient, and the congestion of pulmonary and general circulation may occur. HF is categorized into four

levels by the American College of Cardiology Foundation/American Heart Association guidelines. Only patients at levels III and IV exhibit significant symptoms [1]. Worldwide, more than 23 million patients are affected by HF, causing a major public health problem and huge economic burden [2]. In the USA, the total cost of nursing HF patients is \$31 billion, and this figure is estimated to increase to \$70 billion by 2030 [3]. In addition, the treatment of heart disease comprises the highest health-care costs of low- and middle-income countries.

Echocardiography is often used to diagnose HF in hospitals. This instrument uses ultrasound to measure the stroke volume, end diastolic volume and the ratio between these two

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

quantities, which is also known as the ejection fraction. The general ejection fraction should be between 50% and 70% and less than 40% in the chronic systolic HF. Electrocardiogram (ECG) is another method for detecting CHF. The standard 12-lead ECG remains the most useful instrument in the diagnosis and prognosis of CHF patients. However, the reliability of diagnoses could be further enhanced by signal processing techniques and biomedical analysis [4], [44]. In recent years, many researchers have worked on CHF detection using ECG. For instance, Dhingra *et al.* [5] showed that a longer electrocardiographic QRS duration is associated with CHF. Tereshchenko *et al.* [6] used beat-to-beat QT variability to separate the healthy individuals from HF subjects. Among the recent methods, heart rate variability (HRV) analysis has attracted wide attention for its potential to detect CHF. Nolan *et al.* reported that the standard deviation of the heart rate (SDNN) can effectively predict the risk of mortality for CHF [9]. Binkey *et al.* detected parasympathetic withdrawal by a noninvasive HRV spectral analysis and discovered that this feature is a part of the autonomic nervous disorder feature in CHF patients [8]. Yu and Lee detected CHF with bispectral analysis and a genetic algorithm [10]. Woo *et al.* observed that Poincaré plot analysis can detect labeled sympathetic nerve activation in the CHF patients and identified a link between autonomic nerve change and sudden cardiac death [11]. Peng *et al.* demonstrated a reduction in HR complexity in CHF subjects based on the fractal dimension analysis [12]. Chen *et al.* proposed a dynamic HRV to describe the dynamic fluctuation of HRV over a 24 h period and achieved an over 95% accuracy [15]. However, they admitted that the robustness of HRV-based approaches remains an issue to be addressed as their sensitivity changes according to the clinical condition [7].

However, most of the above studies require long-term nRR interval (RRI) data, including the heart rate testing applications currently being developed for use with mobile devices such as smartphones, which are impossible to obtain in health-care situations outside the hospital. The short-term HRV analysis may be useful in monitoring dynamic changes to cardiac autonomic activity [17]. This method has been used to detect cardiovascular diseases, such as atrial fibrillation (AF), and achieved good results [18]. Thakre *et al.* [19] observed that the lag-response of Poincaré plot indices are related to CHF. Liu *et al.* presented an entropy (Fuzzy-MEn) method to classify the normal and CHF patients [14], and a comparison of the accuracy of entropy arguments on CHF subjects can be found in [13]. Liu and Gao [16] also attempted to detect CHF with short-term RRIs using multiscale entropy analysis based on the RRI signals and a support vector machine classifier. Yoon *et al.* [20] achieved an 84.49% accuracy in CHF detection by observing 16 heartbeats. They extracted three ‘expert features’—normalized root mean squared successive difference (RMSSD), sample entropy (SampEn) and Shannon entropy—and used the threshold values of these variables to detect CHF.

Deep learning (DL) [34], [36] has been applied in various fields, such as image recognition and speech recognition, and achieved remarkable results [21], [33], [35]. In recent years, certain scholars have applied DL to the recognition of ECG signals. Chen *et al.* provided a CHF detection method based on sparse auto-encoder-based DL of short-term RRIs [7]. Potes *et al.* [22] used an ensemble of feature-based and DL-based classifiers for detection of abnormal heart sounds. Hwang *et al.* [23] provided an optimal DL framework for monitoring mental stress using ultra-short-term ECG signals. Pourbabaei *et al.* [24] used deep convolution neural networks to learn the ECG features for screening paroxysmal AF patients. Nevertheless, a work that integrates the learning and expert features used for CHF detection remains inadequate.

In this paper, we propose an ensemble classifier for CHF detection using short-term HRV data and deep neural networks. We first pre-processed and segmented the data as raw input and then explored and implemented the expert features from a large number of studies. We can separate these features into three types: time-domain, frequency-domain and nonlinear features. Secondly, we used the convolutional net architecture (ConvPool-CNN-C) introduced by Springenberg *et al.* [38] to extract the DL features. Finally, we combined the types of features and used them as final input to train an ensemble classifier consisting of several gradient decision boosting tree classifiers. Fig. 1 shows an outline of the method.

The paper is organized as follows. Section 2 presents a detailed description of the proposed method, including the database used, feature extraction, network topology, the basic steps and evaluation methods. Section 3 presents the classification results. Section 4 provides a discussion and describes the limitations of the study, and Section 5 presents the conclusion.

II. MATERIALS AND METHODS

A. DATA

In this study, five open-source databases were used to evaluate the proposed method. For CHF patients, the BIDMC Congestive Heart Failure Database (BIDMC-CHF) [39], [40] and the Congestive Heart Failure RR Interval Database (CHF-RR) [39], which is available on PhysioBank, were used. The BIDMC-CHF dataset included 15 subjects (11 men aged 22–71 years and 4 women aged 54–63 years old) with severe CHF (NYHA class III–IV), whereas the CHF-RR dataset comprised 29 recordings of subjects (aged 34–79 years old) with CHF (NYHA class I–III). For normal subjects, the MIT-BIH Normal Sinus Rhythm (NSR) [39], Normal Sinus Rhythm RR Interval Database (NSR-RR) [39] and Fantasia Database (FD) [53] were used. The NSR included 18 subjects with no significant arrhythmias (5 men aged 26–45 years and 13 women aged 20–50 years), the NSR-RR contained the beat annotation files for 54 subjects with normal sinus rhythm (30 men aged 28.5–76 years old and 24 women aged 58–73 years). The FD comprised 120 min of continuous

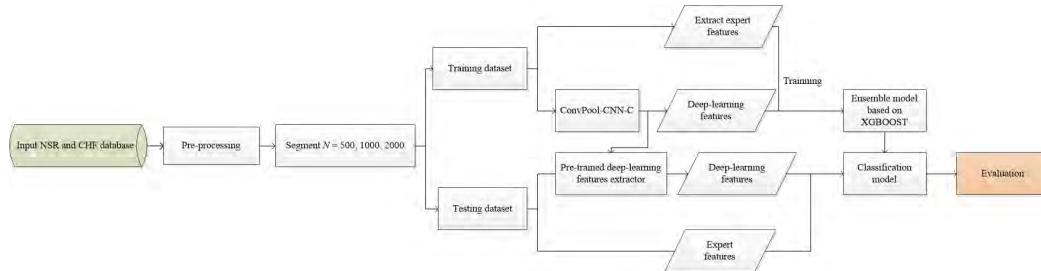


FIGURE 1. Block diagram of the detect procedure.

ECG of 20 young and 20 elderly healthy subjects. In these databases, each beat of the raw ECG signals possessed a label indicating a feature type. The beat labels were firstly produced by a simple slope-sensitive QRS detector and then corrected by two cardiologists.

The pre-processing step in this method included two steps: 1) removing the RRIs greater than 2s for ignoring the interference from the artifacts; 2) removing the RRIs marked as abnormal, such as ectopic beats. To compare our results with those of other works, we segmented the data into 500, 1000 and 2000 samples. Table 1 summarizes the number of signals for the different databases in two classes and those after the above two pre-processing steps. Fig. 2 shows the signals of different feature types for the 500 sample length.

B. FEATURE EXTRACTION

1) TIME DOMAIN FEATURES

Time domain features are widely used in the analysis of HRV, as they are easy to compute and feature important clinical value. We used classical statistical methods to calculate the time-domain characteristics of HRV. Firstly, we calculated the typical clinical indicators, such as the mean RR, minimum RR, std_RR, and median RR. In the next step, we selected another eight time-domain features from the literature: 1) SDNN, RMSSD, NN50, pNN50 and CV [25]; 2) Δ RRImax described by Du *et al.* [26]; and 3) NADev and NADiff proposed by Ghodrati *et al.* [27], which are normalized RRI-based features.

2) FREQUENCY DOMAIN FEATURES

We estimated the power spectrum using the Lomb periodogram method [28]. The Lomb method can avoid the complications and pitfalls of resampling and the replacement of outliers without any drawbacks, making it suitable for the power spectrum estimation of heart rate [29]. We calculated the total power, total power of very-low-frequency range, total power of low-frequency range, total power of high-frequency range, low-frequency in the normalized unit range and high-frequency in the normalized unit range [30].

3) NONLINEAR FEATURES

Nonlinear phenomena often occur in HRV and are determined by the state of the autonomic nervous system, complex hemodynamic interactions and other physiological effects. In this

study, Poincaré plots [19], SampEn [31], Renyi entropy (RyEn) [32] and detrended fluctuation analysis (DFA) [45] were used to extract the nonlinear features.

A Poincaré plot is a graphical representation of the correlation between consecutive RRIs. This plot is a quantitative visual technology that divides the plot shape into functional classes to indicate the degree of heart disease.

Entropy-related methods can be valuable tools for quantifying the regularity of physiological time series and provide important insights into the underlying mechanisms of the cardiovascular system. SampEn [31] features notable consistency with known random series and has been widely used in the analysis of physiological time series, especially the HRV analysis. Cornforth and Jelinek [32] demonstrated that RyEn can identify CHF effectively. The RyEn is a generalized measure and includes the Shannon entropy as a special case:

$$H(\alpha) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^n p_i^\alpha \right) \quad (1)$$

where p_i refers to the probability that a random variable assumes a given value out of n values, and α denotes the order of the entropy measurement.

In recent decades, the DFA introduced by Deng *et al.* [45] has been used as an effective tool for the detection of long-range correlations in time series and has been successfully applied in the study of heart rate dynamics [46], [47].

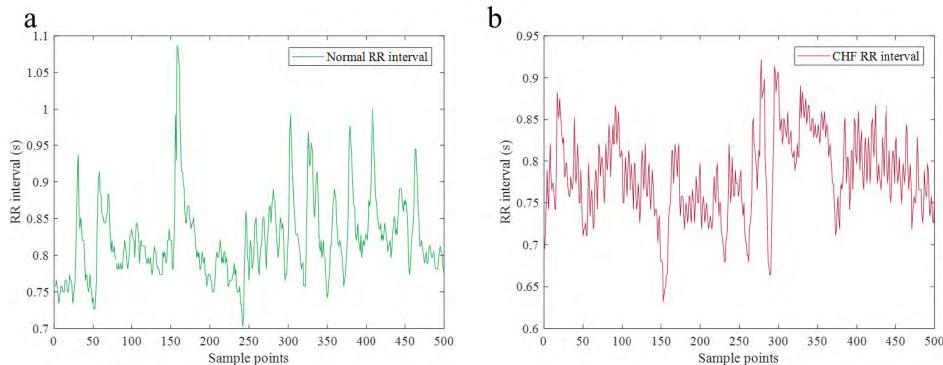
Briefly, the time series to be analyzed (with N samples) was first integrated. Secondly, the integrated time series was divided into boxes of equal length, n . In each box of length n , a least-squares line was fitted to the data. As a result, the y coordinates of the straight-line segments were denoted by $y_n(k)$. In the next step, $y(k)$ can be detrended by subtracting the local trend, $y_n(k)$, in each box. Next, we calculated the mean-square fluctuation of this integrated and detrended time series as follows [45]:

$$F^2(n) = \frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2 \quad (2)$$

This computation was repeated over all time scales (box sizes) to characterize the relationship between $F(n)$, the average fluctuation and the box size, n . Typically, $F(n)$ will increase with box size. A linear relationship on a log-log plot indicates the presence of power law (fractal) scaling. Under such conditions, the fluctuations can be char-

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

Database	Pre-processing	Total Segments		
		500	1000	2000
BIDMC Congestive	None	3225	1612	806
Heart Failure Database	Removing the RR intervals greater than 2s	3214	1607	803
(BIDMC-CHF)	Removing the RR intervals greater than 2s or marked as abnormal beats	3131	1565	782
Congestive Heart	None	6635	3317	1658
Failure RR Interval	Removing the RR intervals greater than 2s	6622	3311	1655
Database (CHF-RR)	Removing the RR intervals greater than 2s or marked as abnormal beats	6338	3169	1584
MIT-BIH Normal	None	3587	1793	896
Sinus Rhythm	Removing the RR intervals greater than 2s	3579	1739	869
Database (NSR)	Removing the RR intervals greater than 2s or marked as abnormal beats	3430	1715	857
Normal Sinus Rhythm	None	11555	5777	2888
RR Interval Database	Removing the RR intervals greater than 2s	11538	5769	2884
(NSR-RR)	Removing the RR intervals greater than 2s or marked as abnormal beats	11431	5715	2857
	None	512	256	128
Fantasia dataset (FD)	Removing the RR intervals greater than 2s	512	256	128
	Removing the RR intervals greater than 2s or marked as abnormal beats	504	252	126

**FIGURE 2.** Signals of different type for 500 samples length. (a) The normal RR interval. (b) The CHF RR interval.

acterized by a scaling exponent: the slope of the line relating $\log F(n)$ to $\log n$.

4) DEEP-LEARNING FEATURES

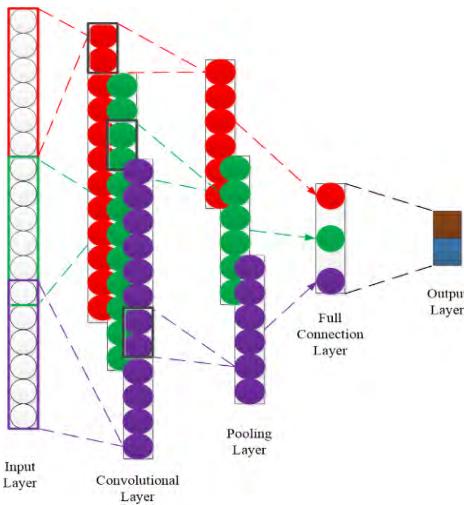
Convolutional neural networks (CNNs) are a kind of deep feed-forward artificial neural network [37]. As they require no pre-processing nor feature extraction, CNNs have been widely used in numerous fields, including physiological signal identification [49]–[51]. To date, CNNs have become the main trend in the field of pattern classification. Current mainstream CNNs generally include the following modules: convolutional layers, pooling layers and fully-connected layers. However, the ConvPool-CNN-C [38] abandons the fully-connected layer of conventional CNNs [37] and replaces it with a global average pooling layer. ConvPool-CNN-C achieves good results on the cifar-10 and cifar-100 databases and exhibits a competitive performance on

the imageNet database. Inspired by [38], a nine-layer ConvPool-CNN-C was trained end-to-end on RRIs and a sequence of labels for each interval. We used a one-dimensional CNN to construct the model, as shown in Fig. 3.

Table 2 shows the network architecture that was used as the base classifier in this study. The loss function used in this study was mean-square error, and the Adam optimizer [48] was used for stochastic optimization. We set the learning parameters to default settings according to the literature (learning rate = 1 and decay = 0) [23].

C. ENSEMBLE LEARNING METHOD

Gradient boosting was first observed by Breiman [43] and later developed by Friedman *et al.* [42]. This procedure is a machine learning method that can integrate a number of weak classifiers as an ensemble model to solve classification problems.

**FIGURE 3.** Typical structure of one-dimension CNN.**TABLE 2.** Network architecture used as the base classifier.

Layer	Parameters
Input Layer	RR interval
Conv1D	filters=64, kernel size=60, padding='same', activation='relu'
Conv1D	filters=64, kernel size=60, padding='same', activation='relu'
MaxPooling1D	
Conv1D	filters=32, kernel size=60, padding='same', activation='relu'
Conv1D	filters=32, kernel size=60, padding='same', activation='relu'
GlobalAveragePooling1D (Deep-learning features extractor)	
Output Layer	Deep-learning Features (32 dim)

Firstly, we considered a training set $\{x, y\}_1^N$, and our goal was to train an approximation $\hat{F}(x)$ to a function $F(x)$ that minimizes the expected value of the loss function that we set as follows (3):

$$\hat{F} = \arg \min_F \mathbb{E}_{x,y}[L(y, F(x))] \quad (3)$$

For the boosting method, $F(x)$ can be obtained by a weighted sum of base learners:

$$F(x) = \sum_{i=1}^M \gamma_i h_i(x) + b \quad (4)$$

where γ_i represents the weight of the base learner h_i , and b corresponds to the sum of bias values in each base learner. According to the empirical risk minimization principle, the method finds $\hat{F}(x)$, which minimizes the average value of the loss function on the training set. The method starts from a model that comprises a constant function $F_0(x)$, and expands it substantially:

$$F_0(x) = \arg \min_\gamma \sum_{i=1}^n L(y_i, \gamma) \quad (5)$$

TABLE 3. Dataset used for comparison.

Database	Database-1 (DB1)	Database-2 (DB2)
BIDMC-CHF	✓	
CHF-RR		✓
NSR	✓	
NSR-RR		✓
FD	✓	

$$F_m(x) = F_{m-1}(x) + \arg \min_{h_m} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + h_m(x_i)) \quad (6)$$

The method applies a steep descent step in the minimization process for simplification, and updates the model according to the following method:

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^n \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)) \quad (7)$$

$$\gamma_m = \arg \min_\gamma \sum_{i=1}^n L(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))) \quad (8)$$

where the derivatives are obtained with respect to function F_i .

For ensemble learning, we used Xgboost software [41] to determine the importance and interpretation of the above features. Xgboost is a scalable end-to-end boosting system based on the gradient boosting framework introduced by Friedman *et al.* [42].

D. STATISTICAL ANALYSIS

As mentioned before, the RR interval was segmented by different length in this study. As a result, there will be multiple RR intervals for one NSR subject or CHF patient, so as the HRV indices value. We calculated their averages values as the individual results, and computed the mean and standard deviation across all subjects. After that, Kolmogorov-Smirnov test was used for testing whether the indices met the normal distribution. If they met the normal distribution, the group *t*-test was used to test the statistical difference between the NSR and CHF groups. If not, non-parametric test was used [16], [54].

E. EVALUATION METHOD

In this study, three indicators were used for testing: sensitivity, specificity and accuracy. The definitions of the indices are as follows:

$$\begin{aligned} \text{Sensitivity : } Se &= \frac{TP}{TP + FN} \\ \text{Specificity : } Sp &= \frac{TN}{TN + FP} \\ \text{Accuracy : } Acc &= \frac{TP + TN}{TP + FN + TN + FP} \end{aligned} \quad (9)$$

TABLE 4. Overview statistical results of all HRV indices for the NSR and CHF groups.

N	HRV indices	NSR	CHF	p-Value	HRV indices	NSR	CHF	p-Value
500	mean_RR	0.87±0.16	0.70±0.11	5×10 ⁻⁷	TP	0.99±0.14	0.95±0.06	8×10 ⁻⁵
	std_RR	0.07±0.33	0.08±0.06	0.1796	vLF	0.90±0.13	0.86±0.05	8×10 ⁻⁵
	min_RR	0.64±0.13	0.56±0.09	6×10 ⁻⁴	LF	0.07±0.01	0.07±0.002	0.0067
	media_RR	0.88±0.15	0.68±0.10	2×10 ⁻⁹	HF	0.02±0.002	0.02±0.00	0.6528
	CV	0.08±0.04	0.11±0.08	0.0015	LF/HF	4.37±0.06	4.32±0.14	0.0015
	ΔRRIMAX	0.53±0.21	0.65±0.31	0.0088	LF.nu	81.38±0.23	81.16±0.52	6×10 ⁻⁴
	NADev	29.33±20.84	20.50±12.17	1×10 ⁻⁵	HF.nu	18.61±0.25	18.83±0.53	6×10 ⁻⁴
	NADiff	17.51±18.82	10.75±11.45	6×10 ⁻⁷	SE	1.35±0.03	0.91±0.43	1×10 ⁻⁹
	SDNN	0.06±0.05	0.10±0.09	1×10 ⁻⁴	SD1	0.04±0.03	0.07±0.06	2×10 ⁻⁴
	RMSSD	0.04±0.03	0.09±0.07	6×10 ⁻⁶	SD2	0.08±0.03	0.08±0.06	0.0252
1000	NN50	95.10±167.63	44.89±60.94	0.0259	RE	6.21±0.25	6.05±0.15	2×10 ⁻⁷
	pNN50	14.72±16.36	8.99±12.21	0.0259	DFA	0.012±0.010	0.017±0.012	9×10 ⁻⁴
	mean_R	0.82±0.11	0.70±0.11	3×10 ⁻⁶	TP	0.78±0.23	0.58±0.03	1×10 ⁻⁷
	std_RR	0.07±0.03	0.08±0.06	0.0544	vLF	0.72±0.21	0.54±0.03	1×10 ⁻⁷
	min_RR	0.57±0.12	0.54±0.10	0.1053	LF	0.05±0.02	0.03±0.00	2×10 ⁻⁸
	media_RR	0.82±0.12	0.68±0.10	9×10 ⁻⁸	HF	0.01±0.004	0.01±0.00	6×10 ⁻⁴
	CV	0.09±0.04	0.12±0.08	0.0163	LF/HF	4.38±0.09	4.34±0.08	0.0048
	ΔRRIMAX	0.59±0.21	0.75±0.31	0.0014	LF.nu	81.36±0.55	81.24±0.30	0.0320
	NADev	47.38±21.03	43.52±23.88	0.1239	HF.nu	18.64±0.55	18.75±0.29	0.0320
	NADiff	25.03±17.61	21.52±22.82	8×10 ⁻⁵	SE	1.28±0.26	0.85±0.40	2×10 ⁻⁴
2000	SDNN	0.06±0.05	0.10±0.08	4×10 ⁻⁴	SD1	0.04±0.03	0.07±0.06	4×10 ⁻⁴
	RMSSD	0.04±0.03	0.09±0.07	2×10 ⁻⁴	SD2	0.09±0.03	0.09±0.06	3×10 ⁻⁴
	NN50	124.47±168.44	89.81±121.43	0.0996	RE	6.51±0.38	6.71±0.16	1×10 ⁻⁶
	pNN50	17.64±18.35	8.99±12.16	0.0016	DFA	0.011±0.010	0.016±0.013	1×10 ⁻³
	mean_RR	0.86±0.14	0.70±0.10	5×10 ⁻⁷	TP	0.23±0.04	0.17±0.07	1×10 ⁻⁵
	std_RR	0.07±0.03	0.08±0.06	0.097	vLF	0.21±0.04	0.15±0.07	1×10 ⁻⁵
	min_RR	0.55±0.13	0.52±0.07	0.141	LF	0.02±0.00	0.02±0.00	0.3135
	media_RR	0.87±0.14	0.69±0.10	5×10 ⁻⁷	HF	0.004±0.00	0.004±0.00	0.2396
	CV	0.09±0.04	0.12±0.08	0.0015	LF/HF	4.35±0.17	4.39±0.29	0.2712
	ΔRRIMAX	0.75±0.21	0.86±0.31	0.0122	LF.nu	80.41±1.61	79.42±2.34	0.0023
4000	NADev	117.31±30.90	93.28±45.57	2×10 ⁻⁴	HF.nu	19.57±1.61	20.57±2.34	0.0023
	NADiff	59.74±31.64	43.15±45.39	1×10 ⁻⁵	SE	1.21±0.26	0.82±0.35	7×10 ⁻⁵
	SDNN	0.06±0.04	0.12±0.08	1×10 ⁻⁴	SD1	0.04±0.03	0.07±0.06	1×10 ⁻⁴
	RMSSD	0.04±0.03	0.09±0.07	1×10 ⁻⁵	SD2	0.09±0.03	0.09±0.07	0.0033
	NN50	287.20±298.69	179.56±241.29	0.0456	RE	7.52±0.07	7.36±0.18	2×10 ⁻⁵
8000	pNN50	14.36±14.94	8.98±12.07	0.0456	DFA	0.012±0.009	0.009±0.011	8×10 ⁻⁵

TABLE 5. Statistical results for paired differences of the first and second segment.

Group	HRV indices	difference \pm SEM	p-Value	HRV indices	difference \pm SEM	p-Value
CHF	mean_RR	0.008 \pm 0.011	0.7311	TP	0.0056 \pm 0.0065	0.9271
	std_RR	0.004 \pm 0.0011	0.9270	vLF	0.0052 \pm 0.0059	0.9241
	min_RR	0.0005 \pm 0.026	0.9761	LF	0.0003 \pm 0.0004	0.9241
	media_RR	0.0049 \pm 0.0086	0.8219	HF	2.92 \times 10 $^{-5}$ \pm 6.61 \times 10 $^{-5}$	0.7994
	CV	0.0016 \pm 0.0012	0.9271	LF/HF	0.0125 \pm 0.00146	0.7462
	Δ RRIMAX	0.0485 \pm 0.0236	0.6563	LF.nu	0.0407 \pm 0.0564	0.7407
	NADev	0.7012 \pm 1.0288	0.9999	HF.nu	0.0407 \pm 0.0564	0.7407
	NADiff	0.0009 \pm 0.1794	0.9899	SE	0.0903 \pm 0.0204	0.7406
	SDNN	0.0024 \pm 0.0026	0.9271	SD1	0.0016 \pm 0.0018	0.9271
	RMSSD	0.0002 \pm 0.0018	0.9961	SD2	0.0070 \pm 0.0018	0.9271
NSR	NN50	4.3448 \pm 5.8621	0.9961	RE	0.0123 \pm 0.0142	0.9962
	pNN50	0.8710 \pm 0.9336	0.9851	DFA	0.0022 \pm 0.0007	0.9271
	mean_R	0.0416 \pm 0.0727	0.6833	TP	0.0218 \pm 0.0452	0.7654
	std_RR	0.0099 \pm 0.0399	0.7654	vLF	0.0214 \pm 0.0452	0.7394
	min_RR	0.0017 \pm 0.1117	0.7654	LF	0.0004 \pm 0.0041	0.9771
	media_RR	0.0429 \pm 0.0725	0.7503	HF	9.70 \times 10 $^{-5}$ \pm 4.64 \times 10 $^{-4}$	0.7654
	CV	0.0029 \pm 0.0511	0.7654	LF/HF	0.0053 \pm 0.3057	0.7394
	Δ RRIMAX	0.0503 \pm 0.3900	0.9771	LF.nu	0.0143 \pm 1.0698	0.7154
	NADev	0.2904 \pm 1.788	0.7654	HF.nu	0.0143 \pm 1.0698	0.7154
	NADiff	0.4167 \pm 0.0491	0.9771	SE	0.1153 \pm 0.6071	0.8230
ALL	SDNN	0.0052 \pm 0.0491	0.8995	SD1	0.0038 \pm 0.0347	0.8995
	RMSSD	0.0023 \pm 0.0449	0.9771	SD2	0.0114 \pm 0.0475	0.8625
	NN50	4.534 \pm 1.5637	0.9998	RE	0.0256 \pm 0.1187	0.9771
	pNN50	1.1492 \pm 9.8624	0.9990	DFA	0.0019 \pm 0.0059	0.8995
	mean_RR	0.0103 \pm 0.1553	0.7430	TP	0.0016 \pm 0.0812	0.6482
	std_RR	0.0025 \pm 0.0477	0.9974	vLF	0.0017 \pm 0.0799	0.6482
	min_RR	0.0047 \pm 0.1699	0.7933	LF	1.35 \times 10 $^{-4}$ \pm 0.0059	0.9567
	media_RR	0.0114 \pm 0.1674	0.4639	HF	9.39 \times 10 $^{-6}$ \pm 7.14 \times 10 $^{-4}$	0.6482
	CV	0.0017 \pm 0.0623	0.5536	LF/HF	0.1085 \pm 1.564	0.8309
	Δ RRIMAX	0.0287 \pm 0.3459	0.8309	LF.nu	0.6376 \pm 12.633	0.8309
	NADev	3.5805 \pm 5.1475	0.5925	HF.nu	0.6376 \pm 12.633	0.8309
	NADiff	0.0724 \pm 41.18	0.9044	SE	0.0147 \pm 0.423	0.7855
	SDNN	0.006 \pm 0.0697	0.8309	SD1	0.0042 \pm 0.0493	0.8309
	RMSSD	0.0021 \pm 0.0462	0.9044	SD2	9.32 \times 10 $^{-4}$ \pm 0.0530	0.8309
	NN50	40.518 \pm 360.54	0.8309	RE	0.0030 \pm 0.100	0.8309
	pNN50	1.9950 \pm 18.10	0.7430	DFA	0.0022 \pm 0.0132	0.9044

*SE means Sample Entropy and RE means Renyi Entropy.

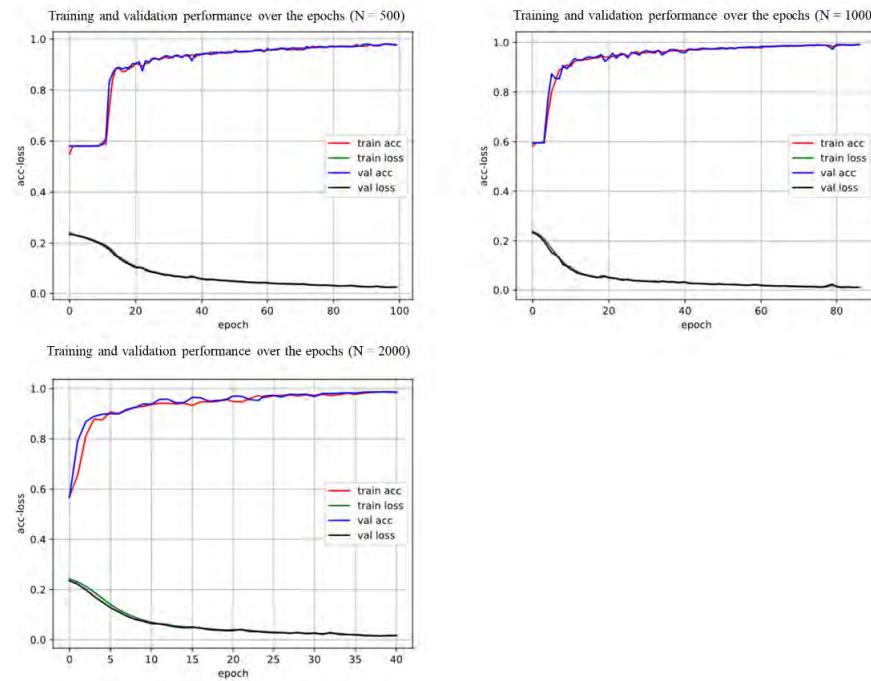
where TP denotes the number of true positives, FN refers to the number of false negatives, FP specifies the number of false positives, and TN stands the number of true negatives.

III. RESULTS

To verify the proposed method, we compared the results of the proposed approach with those of other studies. However,

TABLE 6. The overall performance of the training process.

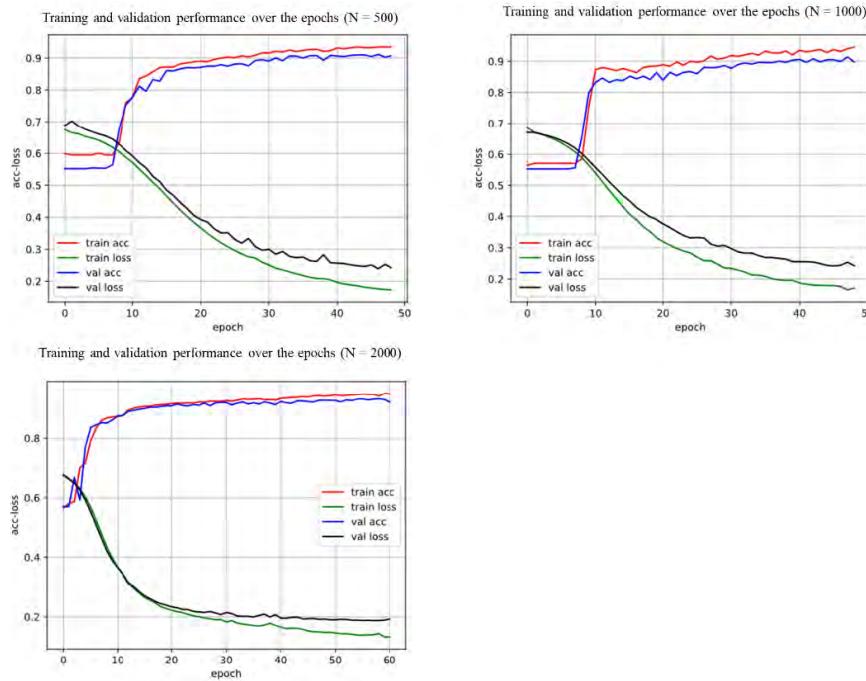
Dataset	Network parameters for each Conv1D layer (filters number, kernel size)	Segment Length		
		500	1000	2000
BIDMC-CHF, NSR, FD (DB1)	(32, 30), (32, 30), (32, 30), (32, 30)	97.57%	99.27%	98.85%
	(32, 60), (32, 60), (32, 30), (32, 30)	99.57%	99.41%	99.57%
	(32, 60), (32, 60), (32, 60), (32, 60)	99.29%	98.48%	97.78%
	(64, 60), (64, 60), (32, 60), (32, 60)	99.96%	99.23%	99.92%
CHF-RR, NSR-RR (DB2)	(64, 120), (64, 120), (32, 60), (32, 60)	99.89%	99.58%	100%
	(32, 30), (32, 30), (32, 30), (32, 30)	84.02%	89.72%	84.40%
	(32, 60), (32, 60), (32, 30), (32, 30)	91.28%	84.35%	86.79%
	(32, 60), (32, 60), (32, 60), (32, 60)	86.82%	88.61%	85.64%
	(64, 60), (64, 60), (32, 60), (32, 60)	90.32%	89.43%	88.07%
	(64, 120), (64, 120), (32, 60), (32, 60)	88.43%	87.31%	88.23%

**FIGURE 4.** Training and validation loss function over the epochs (DB1).**TABLE 7.** The overall performance of the training process.

Method	Dataset	Segment Length	Evaluation		
			Se	Sp	Acc
The proposed method	BIDMC-CHF, NSR, FD (DB1)	500	99.97%	99.95%	99.96%
		1000	99.57%	98.86%	99.23%
		2000	100.00%	99.83%	99.92%
	CHF-RR, NSR-RR (DB2)	500	78.95%	95.89%	90.32%
		1000	75.32%	96.36%	89.93%
		2000	73.12%	95.42%	88.07%

other studies used different datasets to verify their methods. For instance, Kumar *et al.* [44] used the BIDMC-CHF, NSR

and FD databases for CHF detection. Liu and Gao [16] used the NSR-RR and CHF-RR databases, whereas Chen *et al.* [7]

**FIGURE 5. Training and validation loss function over the epochs (DB2).****TABLE 8. Information of the blind testing dataset.**

Database	Subject information (Age, Sex, Number)		Total Segments		
	CHF	Normal	N=500	N=1000	N=2000
Blind Testing subjects for Database-1	(54, F, #11)	(50, F, #19830)			
	(63, M, #13)	(38, F, #19140)	686	339	164
	(61, M, #14)	(34, M, #19093)			
Blind Testing subjects for Database-2	(35, unknown, #224)	(39, M, #049)			
	(66, unknown, #225)	(29, M, #050)			
	(51, unknown, #226)	(40, M, #051)			
	(64, unknown, #227)	(35, M, #054)	2707	1343	662
	(51, unknown, #228)	(64, F, #001)			
	(58, unknown, #229)	(67, F, #003)			

used the 5 min RRI from the same database. Kumar *et al.* [44] used the BIDMC-CHF database, NSR database and FD for CHF detection. Therefore, in this study, to examine the proposed method, we used the same datasets for comparison, as shown in Table 3.

A. STATISTICAL RESULTS

Table 4 shows the overview statistical results (mean \pm standard deviation) of Top-20 effective features for the two groups when setting the segment length as 500, 1000 and 2000, respectively. The indices without statistical difference between the two groups were marked as gray shadows.

Besides that, we also extracted the first two segments from each subject/patient and calculated the difference results (difference \pm standard error of mean) of all HRV indices

with the segment length as 500. As shown in Table 5, there was no significant different between the two RR interval segments in both NSR and CHF groups, as well as in the all groups.

B. TRAINING STAGE

The early stopping method was used in the training stage for preventing overfitting. We hold out 20% of the training data for validation in this stage. The early stopping method stops training when the validation loss has stopped improving. Different network parameters may affect the results. Therefore, we adjusted various kinds of network parameter to reach the best classification results. Since the shortest RR interval length is 500 data points in this study, we set the kernel size as 30, 60 and 120 respectively. The experimental accuracy

TABLE 9. Results and comparison with a Signal Length of 500 Samples.

Method	Classifier	Features	Dataset	Evaluation		
				Se	Sp	Acc
[44]	LS-SVM	Accumulated Fuzzy Entropy and (AFEnt) Accumulated Permutation Entropy (APEnt)	BIDMC-CHF, NSR, FD (DB1)	98.07%	98.33%	98.21%
This method	Ensemble Classifier	Expert features and deep-learning features	BIDMC-CHF, NSR, FD (DB1)	100%	99.84%	99.85%
[7]	DNNs	Sparse-auto-encoder features	CHF-RR, NSR-RR (DB2)	49.09%	86.33%	72.86%
[55]	Inception_V4	Fuzzy GMEn	CHF-RR, NSR-RR (DB2)	81.85% accuracy with 300 length RR interval segments.		
[16]	None	Multiscale entropy of ΔRR	CHF-RR, NSR-RR (DB2)	The area under curve (AUC) is 79.5%.		
This method	Ensemble Classifier	Expert features and deep-learning features	CHF-RR, NSR-RR (DB2)	70.26%	98.44%	83.84%

TABLE 10. Results and comparison with a signal length of 1000 samples.

Method	Classifier	Features	Dataset	Evaluation		
				Se	Sp	Acc
[44]	LS-SVM	Accumulated Fuzzy Entropy and (AFEnt) Accumulated Permutation Entropy (APEnt)	BIDMC-CHF, NSR, FD (DB1)	97.95%	98.07%	98.01%
This method	Ensemble Classifier	Expert features and deep-learning features	BIDMC-CHF, NSR, FD (DB1)	99.37%	100.00	99.41%
[16]	SVM	Multiscale entropy of ΔRR	CHF-RR, NSR-RR (DB2)	86.2%	85.2%	85.5%
This method	Ensemble Classifier	Expert features and deep-learning features	CHF-RR, NSR-RR (DB2)	76.71%	99.22%	87.54%

TABLE 11. Results and comparison with a signal length of 2000 samples.

Method	Classifier	Features	Dataset	Evaluation		
				Se %	Sp %	Acc %
[44]	LS-SVM	Accumulated Fuzzy Entropy and (AFEnt) Accumulated Permutation Entropy (APEnt)	BIDMC-CHF, NSR, FD (DB1)	97.76%	97.67%	97.71%
This method	Ensemble Classifier	Expert features and deep-learning features	BIDMC-CHF, NSR, FD (DB1)	100.00%	98.59%	99.17%
[16]	SVM	Multiscale entropy of ΔRR	CHF-RR, NSR-RR (DB2)	84.4%	86.8%	85.6%
This method	Ensemble Classifier	Expert features and deep-learning features	CHF-RR, NSR-RR (DB2)	73.16%	99.04%	85.71%

of validation dataset under different network parameters are listed in Table 6. By comprehensive consideration, the final parameters are set as Table 2 in Section II.

Fig. 4 and 5 shows the training process details of different training datasets, whereas Table 7 lists the overall performance of the training process.

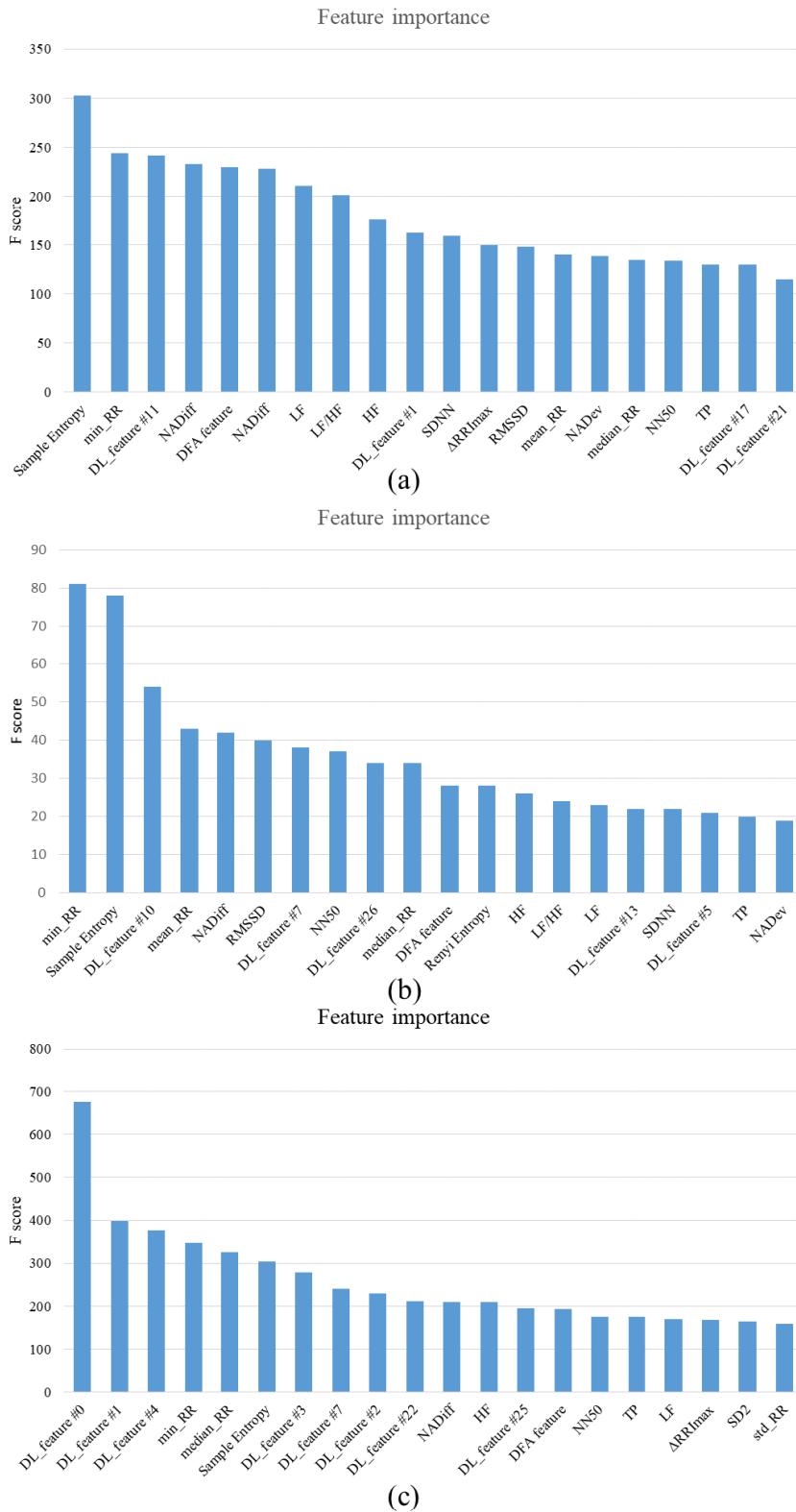


FIGURE 6. Feature importance (DB1), (a) segment length = 500; (b) segment length = 1000; (c) segment length = 2000.

C. BLIND TESTING RESULTS

In practice, the classification system must deal with completely unknown subjects and not with unknown signal

sequences of otherwise known signals, as in the case of cross validation. We have tested the proposed method with blind validation, and the information on the blind validation dataset

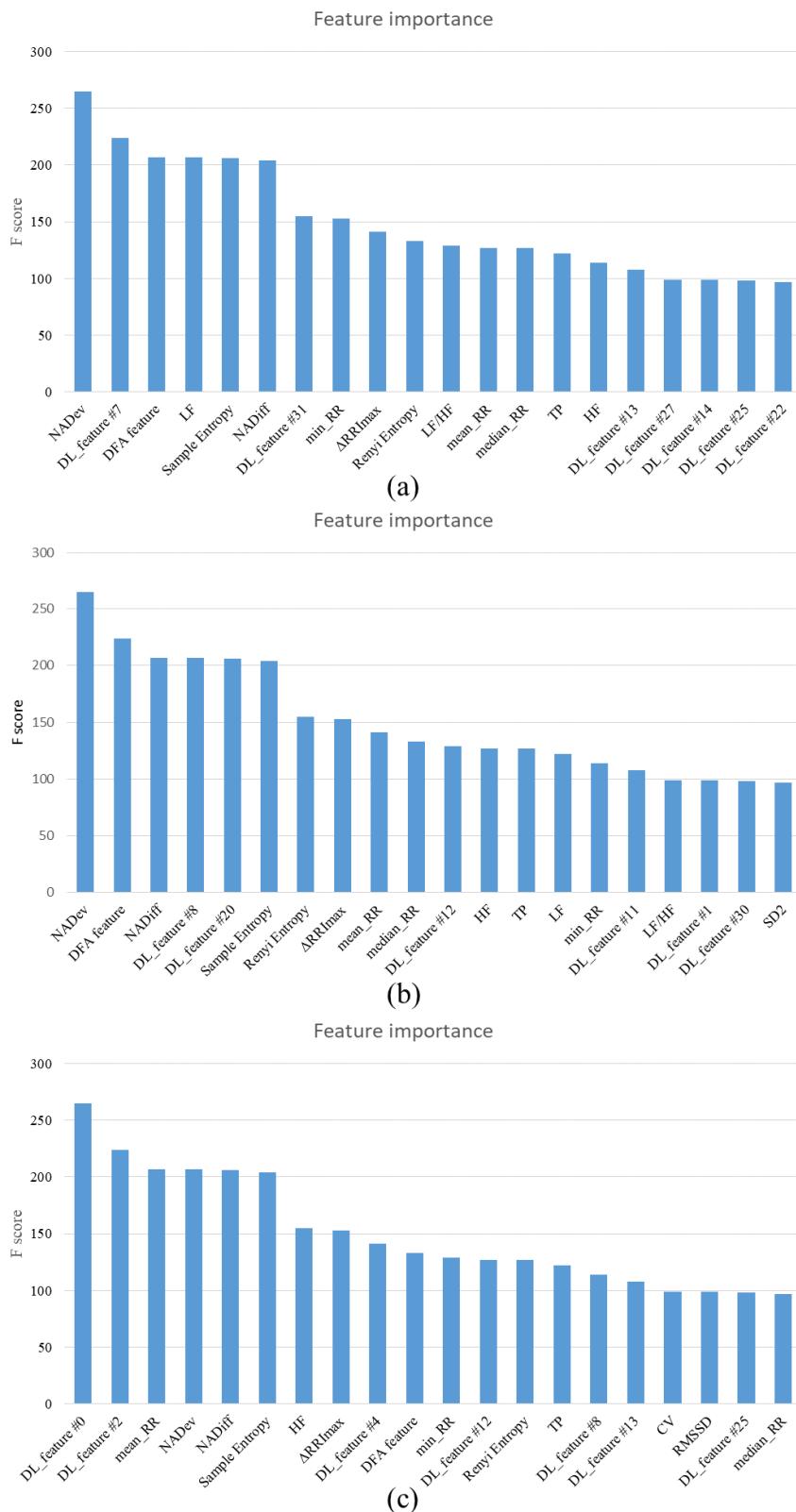


FIGURE 7. Feature importance (DB2), (a) segment length = 500; (b) segment length = 1000; (c) segment length = 2000.

TABLE 12. Rank of features importance (DB1).

Feature Type	Total	Segment Length	Top-5	Top-10	Top-20
Deep-learning features	32 (#0~#31)	500	1	2	4
		1000	1	3	5
		2000	3	7	8
Time domain features	12 (#32~#43)	500	2	3	9
		1000	3	6	8
		2000	2	2	6
Frequency domain features	7 (#44~#50)	500	0	3	5
		1000	0	0	4
		2000	0	0	3
Nonlinear features	5 (#51~#55)	500	2	2	2
		1000	1	1	3
		2000	0	1	3

TABLE 13. Rank of features importance (DB2).

Feature Type	Total	Segment Length	Top-5	Top-10	Top-20
Deep-learning features	32 (#0~#31)	500	1	2	7
		1000	2	2	6
		2000	2	3	7
Time domain features	12 (#32~#43)	500	1	4	7
		1000	2	5	6
		2000	3	5	8
Frequency domain features	7 (#44~#50)	500	1	1	3
		1000	0	0	4
		2000	0	1	2
Nonlinear features	5 (#51~#55)	500	2	3	3
		1000	1	3	4
		2000	0	1	3

are listed in Table 8. Table 9, 10 and 11 provide the results and comparisons for different signal lengths.

The results showed that the proposed ensemble classifier achieves a significant performance on the test set of completely unseen patients.

D. FEATURES IMPORTANCE

We also evaluated the importance of all input features by the average information gain used in each decision trees. In general, importance is represented by a score that indicates how useful or valuable all the features were in the construction of the boosted decision trees. A higher relative importance is achieved when an attribute is used more to make key decisions with a model. We plotted the features ordered by their importance, as shown in Figs. 6 and 7. We also listed the details of feature importance in Tables 12 and 13.

As shown in Table 12 and 13, frequency domain features are less effective than other statistical and deep-learning features, especially for the subjects with less severe CHF (DB2), with a value of 1-2 in the Top-20. Notably, the DL features exert significant effects on the CHF detection.

IV. DISCUSSION

In this study, an ensemble classifier for CHF detection using short-term HRV and deep neural networks was proposed. Five open-source databases were used, and three RR segment length types ($N = 500, 1000, 2000$) were applied to evaluate the proposed method. With blindfold validation (three CHF subjects and three normal subjects), the proposed method achieved 99.85%, 99.41% and 99.17% accuracy on $N = 500, 1000, 2000$ length of RRIs, respectively, using the BIDMC-CHF, NSR and FD databases. With blindfold validation (six

CHF subjects and six normal subjects), the proposed method achieved 83.84%, 87.54% and 85.71% accuracy on $N = 500$, 1000, 2000 length of RRIs, respectively, using the NSR-RR and CHF-RR databases.

A possible explanation for the better performance of our method is that the DL features allow more reliable signal abstraction in the high-dimensional space without human operation. The DL system forms a more abstract high-level representation of attribute classes or features by combining low-level features to discover the distributed feature representations in the data. In addition, we combined the DL features with the expert features, which are highly related to the domain knowledge.

However, this study presents several limitations. Firstly, we disregarded the problem on data imbalance. Table 1 shows the uneven sample sizes of healthy and CHF subjects, especially for the experiments using the CHF-RR and NSR-RR datasets. Secondly, only one deep network structure was used in this study. For ensemble learning, the classification performance may be improved by using more feature extractors with large structural differences.

V. CONCLUSIONS

In summary, this study proposed an ensemble classifier for CHF detection, achieving good classification performance. In our future work, we will attempt to solve the issue on data imbalance and apply more complex network structures, such as recurrent neural networks. We expect the method to be a useful automatic tool to increase the detection rate of patients with CHF.

REFERENCES

- [1] C. W. Yancy *et al.*, “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*, vol. 128, no. 16, pp. 1810–1852, 2012.
- [2] S. M. Dunlay and V. L. Roger, “Understanding the epidemic of heart failure: Past, present, and future,” *Current Heart Failure Rep.*, vol. 11, no. 4, pp. 404–415, 2014.
- [3] P. A. Heidenreich, N. M. Albert, L. A. Allen, D. A. Bluemke, J. Butler, G. C. Fonarow, J. S. Ikonomidis, O. Khavjou, M. A. Konstam, T. M. Maddox, G. Nichol, M. Pham, I. L. Piña, and J. G. Trogdon, “Forecasting the impact of heart failure in the United States: A policy statement from the American heart association,” *Circulat., Heart Failure*, vol. 6, no. 3, pp. 606–619, 2013.
- [4] J. K. Rustad, T. A. Stern, K. A. Hebert, and D. L. Musselman, “Diagnosis and treatment of depression in patients with congestive heart failure: A review of the literature,” *Primary Care Companion CNS Disorders*, vol. 15, no. 4, 2013, Art. no. 13r01511.
- [5] R. Dhingra, M. J. Pencina, T. J. Wang, B.-H. Nam, E. J. Benjamin, D. Levy, M. G. Larson, W. B. Kannel, R. B. D’Agostino, and R. S. Vasan, “Electrocardiographic QRS duration and the risk of congestive heart failure: The Framingham heart study,” *Hypertension*, vol. 47, no. 5, pp. 861–867, 2006.
- [6] L. G. Tereshchenko, I. Cygankiewicz, S. McNitt, R. Vazquez, A. Bayes-Genis, L. Han, S. Sur, J. P. Couderc, R. D. Berger, A. B. de Luna, W. Zareba, “Predictive value of beat-to-beat QT variability index across the continuum of left ventricular dysfunction: Competing risks of noncardiac or cardiovascular death and sudden or nonsudden cardiac death,” *Circulat. Arrhythmia Electrophysiol.*, vol. 5, no. 4, pp. 719–727, 2012.
- [7] W. Chen, G. Liu, S. Su, Q. Jiang, and H. Nguyen, “A CHF detection method based on deep learning with RR intervals,” in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2017, pp. 3369–3372.
- [8] P. F. Binkley, E. Nunziata, G. J. Haas, S. D. Nelson, and R. J. Cody, “Parasympathetic withdrawal is an integral component of autonomic imbalance in congestive heart failure: Demonstration in human subjects and verification in a paced canine model of ventricular failure,” *J. Amer. College Cardiol.*, vol. 18, no. 2, pp. 464–472, 1991.
- [9] J. Nolan, P. D. Batin, R. Andrews, S. J. Lindsay, P. Brooksby, M. Mullen, W. Baig, A. D. Flapan, A. Cowley, R. J. Prescott, J. M. Neilson, K. A. Fox, “Prospective study of heart rate variability and mortality in chronic heart failure: Results of the United Kingdom heart failure evaluation and assessment of risk trial (U.K.-heart),” *Circulation*, vol. 98, no. 15, pp. 1510–1516, 1998.
- [10] S.-N. Yu and M.-Y. Lee, “Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability,” *Comput. Biol. Med.*, vol. 42, no. 8, pp. 816–825, 2012.
- [11] M. A. Woo, W. G. Stevenson, D. K. Moser, and H. R. Middlekauff, “Complex heart rate variability and serum norepinephrine levels in patients with advanced heart failure,” *J. Amer. College Cardiol.*, vol. 23, no. 3, pp. 565–569, 1994.
- [12] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, “Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series,” *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 5, no. 1, pp. 82–87, 1995.
- [13] L. Zhao, S. Wei, C. Zhang, Y. Zhang, X. Jiang, F. Liu, and C. Liu, “Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects,” *Entropy*, vol. 17, no. 12, pp. 6270–6288, 2015.
- [14] C. Liu, K. Li, L. Zhao, F. Liu, D. Zheng, C. Liu, and S. Liu, “Analysis of heart rate variability using fuzzy measure entropy,” *Comput. Biol. Med.*, vol. 43, no. 2, pp. 100–108, 2013.
- [15] W. Chen, L. Zheng, K. Li, Q. Wang, G. Liu, and Q. Jiang, “A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement,” *PLoS One*, vol. 11, no. 11, 2016, Art. no. e0165304.
- [16] C. Liu and R. Gao, “Multiscale entropy analysis of the differential RR interval time series signal and its application in detecting congestive heart failure,” *Entropy*, vol. 19, no. 6, p. 251, 2017.
- [17] A. L. Smith, H. Owen, and K. J. Reynolds, “Heart rate variability indices for very short-term (30 beat) analysis. Part 2: Validation,” *J. Clin. Monitor. Comput.*, vol. 27, no. 5, pp. 577–585, 2013.
- [18] D. E. Lake and J. R. Moorman, “Accurate estimation of entropy in very short physiological time series: The problem of atrial fibrillation detection in implanted ventricular devices,” *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 300, no. 1, p. H319, 2011.
- [19] T. P. Thakre and M. L. Smith, “Loss of lag-response curvilinearity of indices of heart rate variability in congestive heart failure,” *BMC Cardiovascular Disorders*, vol. 6, no. 1, p. 27, 2006.
- [20] K.-H. Yoon, Y. Nam, T. Thap, C. Jeong, N. H. Kim, J. S. Ko, S.-E. Noh, and J. Lee, “Automatic detection of congestive heart failure and atrial fibrillation with short RR interval time series,” *J. Electr. Eng. Technol.*, vol. 12, no. 1, pp. 346–355, 2017.
- [21] J. Schmidhuber, “Deep Learning in neural networks: An overview,” *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2014.
- [22] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, “Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds,” in *Proc. Comput. Cardiol. Conf.*, 2017, pp. 621–624.
- [23] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, and B.-T. Zhang, “Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals,” *Telemedicine e-Health*, vol. 24, no. 10, pp. 753–772, 2018.
- [24] B. Pourbabaei, M. J. Roshtkhari, and K. Khorasani, “Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients,” *IEEE Trans. Syst., Man., Cybern., Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.
- [25] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, and F. Lombardi, “Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology and the North American society of pacing and electrophysiology,” *Eur. Heart J.*, vol. 17, pp. 354–381, Mar. 1996.
- [26] X. Du, N. Rao, M. Qian, D. Liu, J. Li, W. Feng, L. Yin, and X. Chen, “A novel method for real-time atrial fibrillation detection in electrocardiograms using multiple parameters,” *Ann. Noninvasive Electrocardiol.*, vol. 19, no. 3, pp. 217–225, 2014.

- [27] A. Ghodrati, B. Murray, and S. Marinello, "RR interval analysis for detection of Atrial Fibrillation in ECG monitors," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 601–604.
- [28] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophys. Space Sci.*, vol. 39, no. 2, pp. 447–462, Feb. 1976.
- [29] G. B. Moody, "Spectral analysis of heart rate without resampling," in *Proc. Comput. Cardiol. Conf.*, Sep. 1993, pp. 715–718.
- [30] I. Gritti, S. Defendi, C. Mauri, G. Banfi, P. Duca, and G. S. Roi, "Heart rate variability, standard of measurement, physiological interpretation and clinical use in mountain marathon runners during sleep and after acclimatization at 3480 m," *J. Behav. Brain Sci.*, vol. 3, no. 1, pp. 26–48, 2013.
- [31] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol. Heart Circulatory Physiol.*, vol. 278, no. 6, p. H2039, 2000.
- [32] D. J. Cornforth and H. F. Jelinek, "Detection of congestive heart failure using Renyi entropy," in *Proc. Comput. Cardiol. Conf.*, 2017, pp. 669–672.
- [33] Y. Duan, Y. Lv, and F.-Y. Wang, "Travel time prediction with LSTM neural network," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst.*, Nov. 2016, pp. 1053–1058.
- [34] I. Lee, D. Kim, S. Kang, and S. Lee, "Ensemble deep learning for skeleton-based action recognition using temporal sliding LSTM networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1012–1020.
- [35] L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong, and A. Acero, "Recent advances in deep learning for speech research at Microsoft," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8604–8608.
- [36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [37] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [38] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," Dec. 2014, *arXiv:1412.6806*. [Online]. Available: <https://arxiv.org/abs/1412.6806>
- [39] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, p. E215, 2000.
- [40] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *J. Amer. College Cardiol.*, vol. 7, no. 3, pp. 661–670, 1986.
- [41] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
- [42] J. Friedman, T. Hastie, and R. Tibshirani, "Special invited paper. Additive logistic regression: A statistical view of boosting," *Ann. Statist.*, vol. 28, no. 2, pp. 337–374, 2000.
- [43] L. Breiman, "Arcing the edge," Dept. Statist., Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. 486, 1997.
- [44] M. Kumar, R. B. Pachori, and U. R. Acharya, "Use of accumulated entropies for automated detection of congestive heart failure in flexible analytic wavelet transform framework based on short-term HRV signals," *Entropy*, vol. 19, no. 3, p. 92, 2017.
- [45] C.-K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Long-range anticorrelations and non-Gaussian behavior of the heartbeat," *Phys. Rev. Lett.*, vol. 70, no. 9, p. 1343, 1993.
- [46] N. Papasimakis and F. Pallikari, "Correlated and uncorrelated heart rate fluctuations during relaxing visualization," *Europhys. Lett.*, vol. 90, no. 4, pp. 1303–1324, 2010.
- [47] M. Carrara, L. Carozzi, T. J. Moss, M. de Pasquale, S. Cerutti, M. Ferrario, D. E. Lake, and J. R. Moorman, "Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy," *Physiol. Meas.*, vol. 36, no. 9, p. 1873, 2015.
- [48] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, Jul. 2015, pp. 1–13.
- [49] J. Liu, T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, A. J. Kovatich, C. C. Benz, D. A. Levine, A. V. Lee, L. Omberg, D. M. Wolf, C. D. Shriver, and V. Thorsson, "An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics," *Cell*, vol. 173, no. 2, pp. 400–416, 2018.
- [50] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Med. Phys.*, vol. 44, no. 4, pp. 1408–1419, 2017.
- [51] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "ENCASE: An Ensemble CLASsifiEr for ECG classification using expert features and deep neural networks," in *Proc. Comput. Cardiol. Conf.*, 2017, pp. 1–4.
- [52] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [53] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *J. Physiol.-Regulatory Integrative Comparative Physiol.*, vol. 271, no. 4, pp. R1078–R1084, 1996.
- [54] Y. Wang, S. Wei, S. Zhang, Y. Zhang, L. Zhao, C. Liu, and A. Murray, "Comparison of time-domain, frequency-domain and non-linear analysis for distinguishing congestive heart failure patients from normal sinus rhythm subjects," *Biomed. Signal Process. Control*, vol. 42, no. 4, pp. 30–36, 2018.
- [55] Y. Li, Y. Zhang, L. Zhao, Y. Zhang, C. Liu, L. Zhang, L. Zhang, Z. Li, B. Wang, E. Ng, J. Li, and Z. He, "Combining convolutional neural network and distance distribution matrix for identification of congestive heart failure," *IEEE Access*, vol. 6, pp. 39734–39744, 2018.



LUDI WANG received the B.S. degree in automation from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014, where she is currently pursuing the Ph.D. degree in control science and engineering with the School of Automation. Her research interests include the pattern recognition and intelligent systems.



WEI ZHOU received the B.S. degree from the School of Clinical Medicine, Union Medical College, Tsinghua University, and the M.D. degree from the Department of Neuroscience, Uppsala University, Uppsala, Sweden. Her research interests include the obesity and fMRI analysis.



QING CHANG received the M.D. degree from the Department of Clinical Medicine, Natong Medical College, in 2000, and the Ph.D. degree in molecular cell biology from The University of Tokyo, in 2013. He was a Postdoctoral Fellow with the School of Medicine, Shanghai Jiao Tong University, from 2013 to 2015. He is currently an Associate Professor and a Master Supervisor with the Shanghai University of Medicine and Health Sciences. He also serves as an Attending Doctor and a Research Fellow with the Shanghai General Practice Medical Education and Research Center, Jiading District Central Hospital Affiliated Shanghai University of Medicine and Health Sciences. He has published more than 20 articles. His research interests focus on the usage and challenge of innovating technology in general practice medicine, such as sequencing, big data, and AI.



JIANGEN CHEN received the Bachelor of Medicine degree. He is currently the Director of the Jiading Industrial District Community Health Service Center, Shanghai. He has engaged in community comprehensive reform, community health management services, and hospital management for more than 20 years and actively explored the transformation of primary health care service model. Focusing on the reform goals of healthy Jiading, grassroots, contracting services, and whole process management, he has carrying out the construction of family doctor practice platform project, promoting the pilot project of the new “3+X” family doctor service model, focus on the construction of contract service team model, service process reengineering, contract service package customization, and community health platform support, in order to realize the contract service of family doctor. He is a member of the Big Data Grassroots Application Committee, the General Medical Committee of the Shanghai Medical Association, and the Management Committee of the Community Health Service of the Shanghai Community Health Association. He was awarded the Top Ten Managers of China Medical and Health Brand, in 2018.



XIAOGUANG ZHOU received the M.S. degree from the Department of Precision Instrument, Tsinghua University, in 1984, and the Ph.D. degree in engineering from the Tokyo University of Agriculture and Technology, Japan. He was a Visiting Professor with the Tokyo University of Agriculture and Technology, from 2001 to 2002, and a JSPS Researcher with Tokyo University, from 2013 to 2014. He is currently a Min Du scholars Professor and a Doctoral Supervisor with Minjiang University. He also serves as the Director of the Engineering Research Center of Information Networks, Ministry of Education. He has authored over ten books, over 100 articles, and over 16 inventions. His research interests include control theory and its application in engineering, the Internet of Things and automated logistics systems, and mechatronics technology. He is a permanent member of the Chinese Association of Automation/Manufacturing Technology Committee and the China Institute of Communications/Equipment Manufacturing Technical Committee.

• • •