


Received December 29, 2018, accepted January 20, 2019, date of publication January 25, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895230

# Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning

LINA ZHAO<sup>1</sup>, CHENGYU LIU<sup>1</sup> <sup>1</sup>, (Member, IEEE), SHOUSHUI WEI<sup>2</sup>,  
CHANGCHUN LIU<sup>2</sup>, AND JIANQING LI<sup>1</sup>

<sup>1</sup>State Key Laboratory of Bioelectronics, Jiangsu Key Laboratory of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

<sup>2</sup>School of Control Science and Engineering, Shandong University, Jinan 250061, China

Corresponding author: Chengyu Liu (chengyu@seu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 81871444 and Grant 61571113.

**ABSTRACT** Physiological signal variability can offer important insight into cardiovascular activity and clinical cardiovascular diseases. Heart rate variability (HRV) and pulse transit time variability (PTTV) are two important time series variabilities. However, combining HRV and PTTV can enhance the classification accuracy for heart failure which is unknown. In this paper, a simultaneous analysis of HRV and PTTV performed on both normal subjects and heart failure patients, was carried out, aiming to investigate the improvement of HRV-based heart failure detection with the assistant of PTTV analysis. Forty normal subjects and forty heart failure patients were enrolled. Standard limb lead-II electrocardiogram and radial artery pressure waveforms were synchronously recorded. HRV and PTTV analysis were performed on the acquired RR and PTT time series using the standard time- (MEAN, SDNN, and RMSSD), frequency- (LF, HF, and LF/HF), and non-linear (SD1, SD2, sample entropy, and fuzzy measure entropy) domain indices. The results showed that all HRV indices except MEAN ( $P = 0.1$ ) and LF/HF ( $P = 0.9$ ) showed significant differences (all  $P < 0.01$ ) between the two group, while only MEAN in PTTV significantly decreases in heart failure patients ( $P < 0.01$ ). Moreover, when combined the HRV, PTTV indices, and the predicted probabilities generated from the distance distribution matrix-based convolutional neural network models, the highest classification performances were achieved by a support vector machine classifier, outputting a sensitivity of 0.93, a specificity of 0.88, and an accuracy of 0.90. This paper demonstrated the potential of PTTV analysis for the detection of clinical heart failure.

**INDEX TERMS** Pulse transit time variability (PTTV), electrocardiogram (ECG), heart rate variability (HRV), heart failure, cardiovascular time series, entropy.

## I. INTRODUCTION

Heart failure is a typical degeneration of the heart function featured by the reduced ability for the heart to pump blood efficiently [1], and is a common and costly clinical syndrome, associated with significant morbidity and mortality [2]. Timely diagnosis is important to optimize evidence-based treatment opportunities, which can delay mortality and improve symptoms. However, heart failure remains insufficiently diagnosed worldwide, especially in early age [2], [3]. Early and precise diagnosis is thus vital in clinic for the treatment.

Cardiovascular variabilities, typically as heart rate variability (HRV), have given an insight into understanding the

abnormalities of heart failure, and can be used to identify the higher-risk patients [4], [5]. HRV has been used as a risk predictor in patients with heart failure [6]–[8]. Depressed HRV can be observed in heart failure patients [9] and even in early stages of ventricular dysfunction [10]. The level of sympathetic activation is highly related to the severity of the impairment [11]. Heart failure patients usually have a higher sympathetic and a lower parasympathetic activity [7], [8]. Over the past years, various HRV indices from time-, frequency- and non-linear domain were used for heart failure detection. For time-domain analysis, from a prospective study on 433 patients, Nolan *et al.* [12] found that SDNN was the most powerful risk predictor of death

for heart failure disease. For frequency-domain analysis, La Rovere *et al.* [5] reported that the low frequency (LF) component was a powerful predictor of sudden death in heart failure patients. Hadase *et al.* [4] confirmed that the very low frequency (VLF) content was a powerful predictor. Guzzetti *et al.* found significantly lower LF power and lower  $1/f$  slope in heart failure patients compared with controls. Moreover, the patients who died during the follow-up period presented further reduced LF power and steeper  $1/f$  slope than the survivors [13]. Binkley *et al.* [8] also used HRV spectral analysis and found that parasympathetic withdrawal, in addition to the augmentation of sympathetic drive, was an integral component of the autonomic imbalance characteristic for heart failure patients. For non-linear analysis, Woo *et al.* [7] demonstrated that Poincare plot analysis was associated with marked sympathetic activation in heart failure patients and may provide additional prognostic information into autonomic alterations and sudden cardiac death. Mäkikallio *et al.* [14] reported a short-term fractal scaling exponent was the strongest predictor of mortality of heart failure. Poon and Merrill [6] found that the short-term variations of beat-to-beat interval exhibited strongly and consistently chaotic behavior in all healthy subjects but not in heart failure patients [6]. Peng *et al.* [15] used detrended fluctuation analysis (DFA) method and confirmed a reduction in time series complexity in heart failure patients. Liu *et al.* [16] reported a decrease of approximate entropy (ApEn) values in heart failure group. Costa *et al.* [17] used the multiscale entropy for classifying heart failure patients and healthy subjects, and achieved best discrimination at the scale 5. Various machine learning methods were also proposed to diagnose patients suffering from heart failure based on HRV. Typical studies included: Isler and Kuntalp [18] proposed a k-nearest neighbor classifier (KNN) and wavelet entropy-based model. Jovic and Bogunovic [19] utilized random forest and combinations of linear and non-linear features of HRV. Pecchia *et al.* [20] designed a classifier based on regression tree and the selected HRV features. Wang *et al.* [21] used a support vector machine (SVM) method combined with several HRV features. Li *et al.* [22] proposed a new method combining convolutional neural network (CNN) and distance distribution matrix (DDM) for identification of heart failure patients.

Different cardiovascular variabilities attribute to various autonomic control mechanisms for cardiovascular system [23]. HRV reflects the variability of RR intervals between successive sinus heartbeats, which results from the impact of autonomic nervous system tone on cardiovascular system [5]. The time interval between ventricular electrical activity and the arrival of a peripheral pulse wave is commonly defined as pulse transit time (PTT) and its variability, i.e., pulse transit time variability (PTTV), can provide an insight for understanding the peripheral circulation [24]–[26]. PTT is mainly mediated by the peripheral sympathetic activity. Thus, the hypothesis in the current study is, the combination analysis of HRV and PTTV can provide a further insight for understanding the level of peripheral

sympathetic activation for heart failure patient, and can help on enhancing the detection accuracy for heart failure. To test this hypothesis, a simultaneous analysis of HRV and PTTV, performed on both normal subjects and heart failure patients, was carried out, aiming to investigate the improvement of HRV-based heart failure detection with the assistant of PTTV analysis.

## II. METHOD

### A. DATABASE

In this study, 40 normal subjects and 40 heart failure patients were enrolled, with the matched age and sex, aged between 30 and 75 years. All subjects gave the informed consent and confirmed that they had not have participated in any other clinical trials in the previous three months. The study obtained full approval of the Clinical Ethics Committee of the Qilu Hospital. The investigation conformed with the principles of expressed in the revised guidelines of Declaration of Helsinki – the fifth revision in the Edinburgh 2000 [27].

Heart failure patients were in classes II-III of the New York Heart Association (NYHA) with functional classification confirmed by an ultrasonic cardiogram (UCG) test. Left ventricular ejection fractions (LVEF) from three cardiac cycles were measured by a cardiologist, and their average value was used as the reference LVEF for that subject. The LVEF from the heart failure patients was less than 50% in this study. The normal subjects had normal UCG and ECG, and normal LVEF in the range of 50-80%. The subject demographic information is given in Table 1.

**TABLE 1. Physiological variables from 40 normal subjects and 40 heart failure patients. Data are expressed as number or mean  $\pm$  standard deviation (SD).**

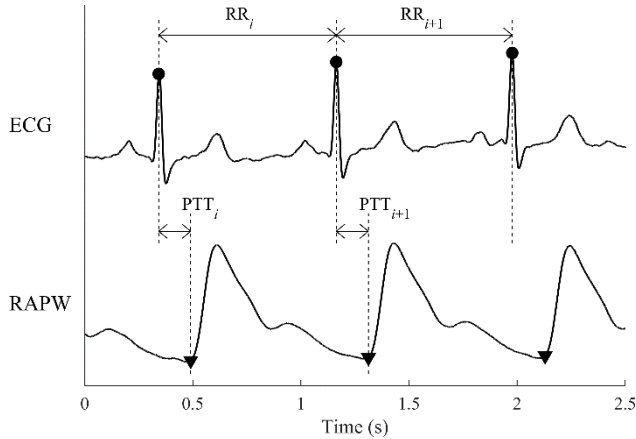
| Variables   | Normal subjects | Heart failure patients | <i>P</i> -values |
|-------------|-----------------|------------------------|------------------|
| No.         | 40              | 40                     | --               |
| Men         | 18              | 21                     | 0.2              |
| Age (year)  | 59 $\pm$ 9      | 59 $\pm$ 8             | 0.4              |
| Height (cm) | 166 $\pm$ 8     | 167 $\pm$ 9            | 0.9              |
| Weight (kg) | 65 $\pm$ 9      | 67 $\pm$ 8             | 0.6              |
| SBP (mmHg)  | 119 $\pm$ 12    | 122 $\pm$ 12           | 0.5              |
| DBP (mmHg)  | 69 $\pm$ 10     | 71 $\pm$ 9             | 0.7              |
| LVEF (%)    | 65 $\pm$ 4      | 36 $\pm$ 8             | <0.01            |

### B. EXPERIMENTAL PROCEDURE

All measurements were undertaken in a quiet, temperature controlled clinical measurement room ( $25 \pm 3^\circ\text{C}$ ) at Qilu Hospital. Before signal recording, each subject lay supine on a measurement bed for a 10 min rest period. Lead-II ECG and radial artery pressure waveforms (RAPW) were acquired with a sample rate of 1,000 Hz and a 5-10 min signal length. Systolic and diastolic blood pressures (SBP and DBP) were manually recorded at the end of the measurement (shown in Table 1).

**C. CONSTRUCTION OF RR AND PTT TIME SERIES**

Figure 1 shows synchronously recorded ECG and RAPW signals. Baselines (0-0.05 Hz) were removed from the original signals using a high-pass filter. R-wave peaks in ECG [28] and the start points of pulse (pulse feet) [29] were detected. RR interval time series were obtained from the adjacent R-wave peaks and PTT interval time series were obtained from the R wave peaks to the feet of the corresponding pulse.



**FIGURE 1.** Synchronously recorded ECG and RAPW signals. The detected R-wave peaks are denoted as “●” and the feet of RAPW are denoted as “▼”. The RR interval is the interval between two adjacent R-wave peaks. The heart-radial PTT is the interval from the R-wave peak to the foot of RAPW signal.

**D. VARIABILITY FEATURES CALCULATION**

HRV and PTTV were quantified by calculating the time-, frequency- and non-linear domain features. For a RR or PTT time series  $x(i)$  ( $1 \leq i \leq N$ , in this study, the length of both RR and PTT time series was set as constant value  $N = 300$ ). The detailed descriptions of the variability features were summarized as follows.

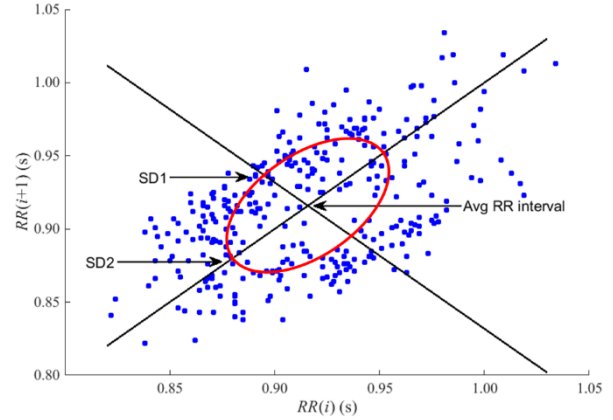
Three time-domain features were defined: the mean value of time series (MEAN), the standard deviation of time series (SDNN) and the square root of the mean of the sum of the squares of differences between adjacent intervals of time series (RMSSD):

$$MEAN = \frac{1}{N} \sum_{i=1}^N x(i) \tag{1}$$

$$SDNN = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x(i) - MEAN)^2} \tag{2}$$

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x(i+1) - x(i))^2} \tag{3}$$

Two spectral components were detected: low frequency power (LF, 0.04-0.15 Hz) and high frequency power (HF, 0.15-0.4 Hz). LF, HF and the ratio between LF and HF (LF/HF) were used as the frequency-domain features.



**FIGURE 2.** Demonstration of the calculation progress for features of SD1 and SD2. Avg means average.

Non-linear features were from several aspects. First, Poincare plot was drawn for the  $x(i)$  (take RR time series for example). In this plot, the length of the front RR interval in two adjacent RR intervals was taken as the horizontal coordinate and the latter as the vertical coordinate (shown in Figure 2). The standard deviation of the distances of the RR intervals to the line  $y = -x + 2RR_{mean}$  was defined as SD1 and to the line  $y = x$  was defined as SD2 [30]. SD1 was related to the fast beat-to-beat variability while SD2 described the longer-term variability in RR time series. Then, entropy-based measures were employed. Entropy measures provide a useful tool for quantifying the regularity of cardiovascular signals. Both sample entropy (SampEn) [31] and the newly developed fuzzy measure entropy (FuzzyME) [32], [33] were used. The detailed calculation processes can be found in [31] and [32] respectively.

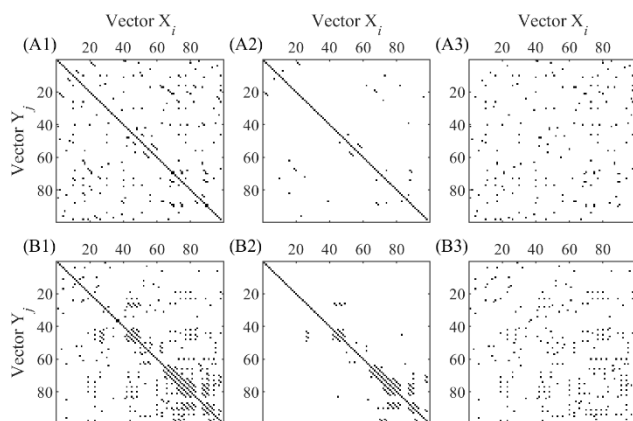
**E. CONSTRUCTION OF DISTANCE DISTRIBUTION MATRIX (DDM)**

Construction of DDM is an essential step for entropy calculation. As an intermediate step, entropy needs to determine the similarity degree between any two vectors  $X_i^m$  and  $Y_j^m$  at both embedding dimension  $m$  and  $m + 1$  respectively, by calculating DDM. The difference between normal and heart failure subjects can be depicted and observed by DDM, thus it is a desirable input for CNN as it reveals the features of times series in the manner of entropy analysis but contains richer information than a simple single entropy value result. We first revealed this DDM method in [34] and then used it for the HRV-based heart failure study [22]. In this study, we employed DDM as a non-linear analysis method for HRV and PTTV. Here, we introduced the construction methods for three DDM types.

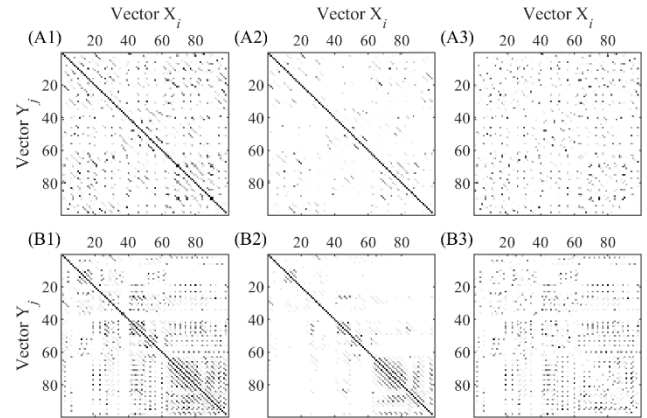
For SampEn, the distance between  $X_i^m$  and  $Y_j^m$  is defined as  $d_{i,j}^m = |x(i+k) - x(j+k)|$ . If the distance is within the threshold parameter  $r$ , the similarity degree between the two vectors is 1; if the distance is beyond the threshold parameter  $r$ , the similarity degree is 0. There is absolutely

a 0 or 1 determination. Unlike the 0 or 1 discrete determination for vector similarity degree in SampEn, FuzzyMEN outputs continuous numerical values between 0 and 1 for vector similarity degree, by converting the absolute distance of  $d_{i,j}^m = |x(i+k) - x(j+k)|$  by a fuzzy exponential function  $u(d_{i,j}^m, n, r) = \exp(-\left(d_{i,j}^m\right)^n / r)$ . FuzzyMEN includes two entropy information from both global vector similarity degree (denoted as FuzzyGMEN), and local vector similarity degree (denoted as FuzzyLMEN). For each of the three entropy methods (SampEn, FuzzyGMEN and FuzzyLMEN), we generated the DDM images.

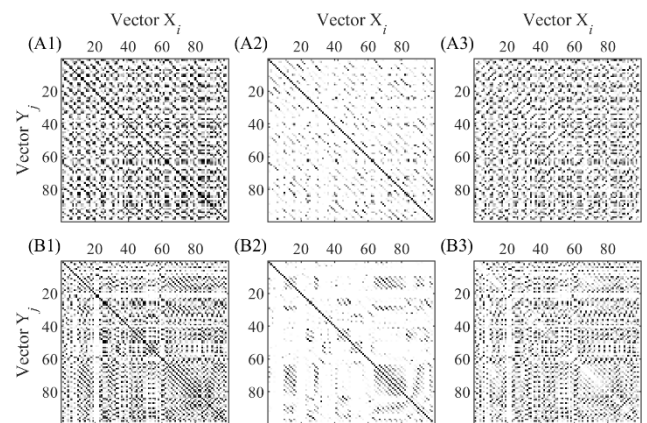
DDM was firstly generated at the setting of embedding dimension  $m$  and  $m + 1$ . Then the difference matrix of these two DDMs, also a DDM, was calculated as the image input for the following CNN classifier. Since each RR or PTT time series has a length of  $N = 300$ , we segmented RR or PTT time series using a fixed window of 100 points with a 95% overlap for DDM generation. Thus, each RR or PTT time series generated 41 difference matrix DDMs as the input of CNN model. Figures 3-5 demonstrate the DDM examples for RR time series generated by SampEn, FuzzyGMEN and FuzzyLMEN respectively. Figures 6-8 demonstrate the DDM examples for PTT time series. In each figure, the upper panels show the results from a normal subject, and the lower panels show the results from a heart failure patient.  $X_i$  and  $Y_j$  are vectors in the time series, at the embedding dimension  $m$  and  $m + 1$ , as well as their difference. Dark colored areas indicated the large similarity degree and vice versa. Number of similar vectors (i.e., matching vectors) decreased when the embedding dimension changes from  $m$  to  $m + 1$ . Fixed parameters were set as: embedding dimension  $m = 2$  and tolerate threshold  $r$  equals 0.1 times of the standard deviation of  $x(i)$ . The reason of using a relatively small tolerate threshold value is that the length of RR or PTT time series is relatively short in this study. Large threshold  $r$  values can lead to the invalid entropy values for short time series as recommended in [35].



**FIGURE 3.** DDM examples for RR time series generated by SampEn: (A1-A3) from a normal subject: (A1)  $m = 2$ , (A2)  $m = 3$ , (A3) the difference of (A1) and (A2); (B1-B3) from a heart failure patient: (B1)  $m = 2$ , (B2)  $m = 3$ , (B3) the difference of (B1) and (B2).



**FIGURE 4.** DDM examples for RR time series generated by FuzzyGMEN: (A1-A3) from a normal subject: (A1)  $m = 2$ , (A2)  $m = 3$ , (A3) the difference of (A1) and (A2); (B1-B3) from a heart failure patient: (B1)  $m = 2$ , (B2)  $m = 3$ , (B3) the difference of (B1) and (B2).

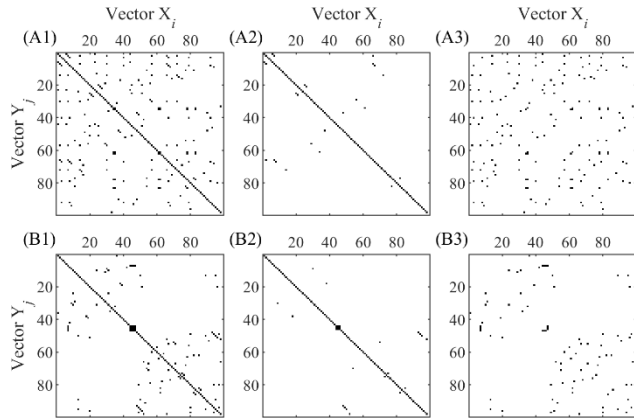


**FIGURE 5.** DDM examples for RR time series generated by FuzzyLMEN: (A1-A3) from a normal subject: (A1)  $m = 2$ , (A2)  $m = 3$ , (A3) the difference of (A1) and (A2); (B1-B3) from a heart failure patient: (B1)  $m = 2$ , (B2)  $m = 3$ , (B3) the difference of (B1) and (B2).

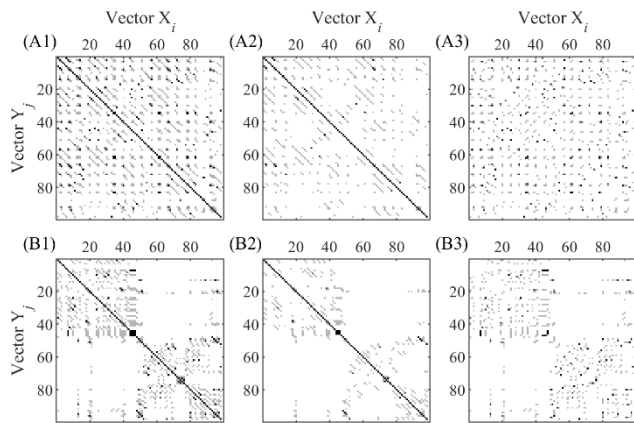
In addition, similarity weight  $n$  was set as 2 for FuzzyGMEN and 3 for FuzzyLMEN.

## F. CNN CLASSIFIER FOR DDM IMAGES

CNN has become a popular method for feature extraction and classification without requiring pre-processing of raw signal. In this study, the difference matrix DDMs (size  $98 \times 98$ ) between embedding dimensions 2 and 3 were used as the input of CNN classifiers. CNN was implemented using the Neural Network Toolbox in Matlab R2017a. A 10-layer network structure was developed as shown in Figure 9, which contained 3 convolution layers (C1 5 filters with kernel size  $5 \times 5$ , C2 10 filters with kernel size  $8 \times 8$  and C3 20 filters with kernel size  $7 \times 7$  respectively), 3 max pooling layers, a flatten layer and a fully connected layer, as well as the input and output layers. The convolution layer extracted various local features of the previous layer by convolution operation. The pooling layer combined similar features and made the feature robust to noise (using Max pooling with kernel size



**FIGURE 6.** DDM examples for PTT time series generated by SampEn: (A1-A3) from a normal subject: (A1)  $m = 2$ , (A2)  $m = 3$ , (A3) the difference of (A1) and (A2); (B1-B3) from a heart failure patient: (B1)  $m = 2$ , (B2)  $m = 3$ , (B3) the difference of (B1) and (B2).



**FIGURE 7.** DDM examples for PTT time series generated by FuzzyGMEn: (A1-A3) from a normal subject: (A1)  $m = 2$ , (A2)  $m = 3$ , (A3) the difference of (A1) and (A2); (B1-B3) from a heart failure patient: (B1)  $m = 2$ , (B2)  $m = 3$ , (B3) the difference of (B1) and (B2).

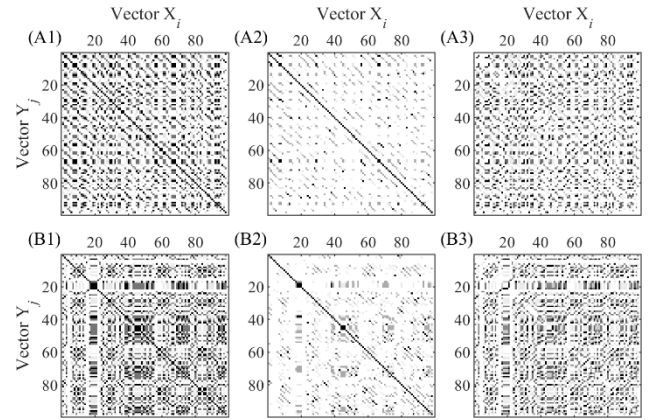
$2 \times 2$  for all three pooling layers). The Flatten layer was used to “flatten” the input, that is, to make multidimensional ( $7 \times 7$  here) input one-dimensional ( $49 \times 1$ ). The fully connected layer combined the various local features extracted in the previous stage, and finally obtained the posterior probability of each category ( $2 \times 1$ ) through the output layer [36]. Figure 9 illustrates the architecture of the implemented CNN model and its detailed components for each layer. Learning rate was set as 0.001 and the number of epochs was set as 20.

### G. CLASSIFICATION AND EVALUATION

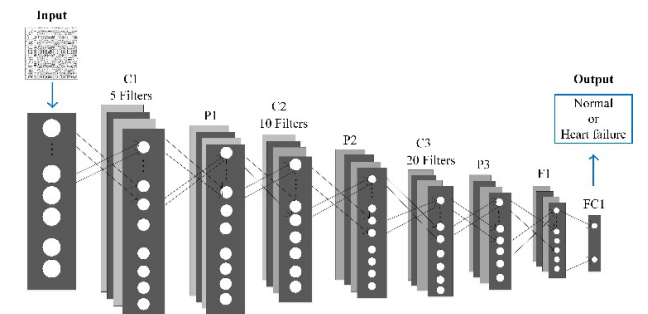
In this study, we performed classification of normal and heart failure subjects using three scenarios.

Scenario I: classification by only CNN-based method (DDMs as the input images).

Scenario II: classification by SVM classifier using HRV information only. The HRV indices includes the time-, frequency- and non-linear domain indices mentioned above. In addition, CNN model can generate predicted probabilities



**FIGURE 8.** DDM examples for PTT time series generated by FuzzyLMEn: (A1-A3) from a normal subject: (A1)  $m = 2$ , (A2)  $m = 3$ , (A3) the difference of (A1) and (A2); (B1-B3) from a heart failure patient: (B1)  $m = 2$ , (B2)  $m = 3$ , (B3) the difference of (B1) and (B2).



**FIGURE 9.** The architecture of the CNN network.

for classifying normal or heart failure. Thus, the predicted probabilities generated from inputting different DDM images (SampEn, FuzzyGMEn and FuzzyLMEn respectively) for RR time series were also used as the SVM input features. Since there are 41 predicted probabilities for each DDM type for each subject, the average results were used as the individual feature.

Scenario III: classification by SVM classifier using the combination of HRV and PTTV information. The HRV and PTTV indices from time-, frequency- and non-linear domain, as well as the predicted probabilities generated from the six DDM images (three for RR and three for PTT time series), were used as the SVM inputs.

Here the open access libsvm software package was used to learn the SVM models [37]. The leave-one-out (LOO) method was used to train and test the SVM classifier. The LOO-based validation was performed considering all subjects in the database and was performed with 80 iterations in this study, such that in each iteration the classifier is trained with 79 subjects and tested on the remaining one sample, to generate the classification result for normal and heart failure classification. It is a total patient-independent test. The Gaussian kernel was used in SVM.  $C$  and  $\gamma$  were optimized using a grid search method with the search range over  $C$  (from 0.1 to 100) and  $\gamma$  (from 1 to 50).

Sensitivity (Se), specificity (Sp) and Accuracy (Acc) defined in (4-6), were used for the evaluation, where true positive (TP) denotes the number of heart failure patients correctly classified as heart failure, false positive (FP) refers to the number of normal subjects incorrectly classified as heart failure, true negative (TN) associates with the number of normal subjects correctly classified as normal, and false negative (FN) refers to the number of heart failure patients incorrectly classified as normal.

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

$$Sp = \frac{TN}{TN + FP} \quad (5)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

### III. RESULTS

#### A. STATISTICAL RESULTS OF INDICES

Mean  $\pm$  SD of all indices were obtained across all subjects (40 normal subjects and 40 heart failure patients). Student's *t*-test was used to compare the statistical differences between two groups. Statistical significance was set a priori at  $P < 0.05$ .

**TABLE 2.** HRV indices from 40 normal subjects and 40 heart failure patients. Data are expressed as mean  $\pm$  SD.

| Variables             | Normal          | Heart failure   | <i>P</i> -values |
|-----------------------|-----------------|-----------------|------------------|
| MEAN (ms)             | 910 $\pm$ 96    | 866 $\pm$ 138   | 0.1              |
| SDNN (ms)             | 37 $\pm$ 16     | 21 $\pm$ 15     | <0.01            |
| RMSSD (ms)            | 34 $\pm$ 21     | 15 $\pm$ 10     | <0.01            |
| LF (ms <sup>2</sup> ) | 355 $\pm$ 447   | 85 $\pm$ 130    | <0.01            |
| HF (ms <sup>2</sup> ) | 576 $\pm$ 676   | 107 $\pm$ 145   | <0.01            |
| LF/HF                 | 1.05 $\pm$ 1.35 | 1.10 $\pm$ 0.89 | 0.9              |
| SD1 (ms)              | 19 $\pm$ 8      | 12 $\pm$ 6      | <0.01            |
| SD2 (ms)              | 26 $\pm$ 13     | 17 $\pm$ 5      | <0.01            |
| SampEn                | 1.68 $\pm$ 0.18 | 1.50 $\pm$ 0.27 | <0.01            |
| FuzzyMEn              | 1.71 $\pm$ 0.27 | 1.31 $\pm$ 0.44 | <0.01            |

Table 2 gives the overall means and SDs of HRV indices (i.e., MEAN, SDNN, RMSSD, LF, HF, LF/HF, SD1, SD2, SampEn and FuzzyMEn) from the two groups. As shown in Table 2, there were no significant differences in MEAN ( $P = 0.1$ ) and LF/HF ( $P = 0.9$ ) between two groups. However, there were significant differences in other eight indices (all  $P < 0.01$ ). For time-domain indices, SDNN in heart failure group was significantly lower by 16 ms (37  $\pm$  16 vs 21  $\pm$  15 ms) and RMSSD was significantly lower by 19 ms (34  $\pm$  21 vs 15  $\pm$  10 ms). For frequency-domain indices, LF in heart failure group was significantly lower by 270 ms<sup>2</sup> (355  $\pm$  447 vs 85  $\pm$  130 ms<sup>2</sup>) and HF was significantly lower by 469 ms<sup>2</sup> (576  $\pm$  676 vs 107  $\pm$  145 ms<sup>2</sup>). For non-linear indices, SD1 in heart failure group was significantly lower by 7 ms (19  $\pm$  8 vs 12  $\pm$  6 ms) and SD2 was significantly lower by 9 ms (26  $\pm$  13 vs 17  $\pm$  5 ms). SampEn in heart failure group was significantly lower by 0.18 (1.68  $\pm$  0.18 vs 1.50  $\pm$  0.27) and FuzzyMEn was significantly lower by 0.40 (1.71  $\pm$  0.27 vs 1.31  $\pm$  0.44).

**TABLE 3.** PTTV indices from 40 normal subjects and 40 heart failure patients. Data are expressed as mean  $\pm$  SD.

| Variables             | Normal          | Heart failure   | <i>P</i> -values |
|-----------------------|-----------------|-----------------|------------------|
| MEAN (ms)             | 125 $\pm$ 13    | 115 $\pm$ 15    | <0.01            |
| SDNN (ms)             | 2.6 $\pm$ 0.7   | 2.8 $\pm$ 1.3   | 0.3              |
| RMSSD (ms)            | 3.3 $\pm$ 1.1   | 3.3 $\pm$ 1.8   | 1.0              |
| LF (ms <sup>2</sup> ) | 0.54 $\pm$ 0.47 | 0.76 $\pm$ 1.05 | 0.3              |
| HF (ms <sup>2</sup> ) | 4.96 $\pm$ 3.66 | 5.21 $\pm$ 6.92 | 0.8              |
| LF/HF                 | 0.14 $\pm$ 0.11 | 0.21 $\pm$ 0.27 | 0.2              |
| SD1 (ms)              | 1.8 $\pm$ 0.4   | 2.7 $\pm$ 0.6   | 0.3              |
| SD2 (ms)              | 2.3 $\pm$ 0.6   | 3.5 $\pm$ 1.0   | 0.4              |
| SampEn                | 1.72 $\pm$ 0.23 | 1.70 $\pm$ 0.45 | 0.8              |
| FuzzyMEn              | 1.91 $\pm$ 0.16 | 1.88 $\pm$ 0.32 | 0.4              |

Table 3 gives the overall means and SDs of PTTV indices from the two groups. As shown in Table 3, only MEAN had significant difference between the two groups (normal group 125  $\pm$  13 vs heart failure group 115  $\pm$  15 ms,  $P < 0.01$ ). However, the other nine indices had no significant differences between the two groups (all  $P > 0.05$ ).

#### B. CLASSIFICATION RESULTS FOR SCENARIO I

Table 4 presents the classification results for Scenario I test, where classification was only based on CNN classifier using DDMs as input. For RR time series, DDMs generated from FuzzyGMEn reported the highest accuracy of 0.83, while DDMs generated from both SampEn and FuzzyLMEn reported accuracies of 0.79. DDMs from PTT time series reported slightly lower accuracies compared with those from RR time series. For PTT time series, DDMs generated from FuzzyLMEn reported the highest accuracy of 0.79. In addition, DDMs from RR time series reported higher sensitivity while DDMs from PTT time series reported higher specificity.

**TABLE 4.** Results for classification of health subjects and heart failure patients using DDM-based CNN classifier.

| DDM                    | TP | FN | FP | TN | Se   | Sp   | Acc  |
|------------------------|----|----|----|----|------|------|------|
| <i>RR time series</i>  |    |    |    |    |      |      |      |
| SampEn                 | 32 | 8  | 9  | 31 | 0.80 | 0.78 | 0.79 |
| FuzzyGMEn              | 34 | 6  | 8  | 32 | 0.85 | 0.80 | 0.83 |
| FuzzyLMEn              | 33 | 7  | 10 | 30 | 0.83 | 0.75 | 0.79 |
| <i>PTT time series</i> |    |    |    |    |      |      |      |
| SampEn                 | 29 | 11 | 9  | 31 | 0.73 | 0.78 | 0.75 |
| FuzzyGMEn              | 30 | 10 | 9  | 31 | 0.75 | 0.78 | 0.76 |
| FuzzyLMEn              | 31 | 9  | 8  | 32 | 0.78 | 0.80 | 0.79 |

#### C. CLASSIFICATION RESULTS FOR SCENARIO II

Table 5 presents the classification results for Scenario II test, where the results were from analysis of only HRV indices, and from both HRV indices and the predicted probabilities from DDM-based CNN models. Using only HRV indices, SVM classifier generated a sensitivity of 0.83 and a specificity of 0.80, resulting in an accuracy of 0.81. However, when combined with the predicted probabilities generated from the DDM-based CNN model, performances of the SVM

**TABLE 5. Results for classification of health subjects and heart failure patients using HRV+DDM-based SVM classifier.**

| Feature  | TP | FN | FP | TN | Se   | Sp   | Acc  |
|----------|----|----|----|----|------|------|------|
| Only HRV | 33 | 7  | 8  | 32 | 0.83 | 0.80 | 0.81 |
| HRV+DDM  | 36 | 4  | 6  | 34 | 0.90 | 0.85 | 0.88 |

**TABLE 6. Results for classification of health subjects and heart failure patients using HRV+PTTV+DDM-based SVM classifier.**

| Feature      | TP | FN | FP | TN | Se   | Sp   | Acc  |
|--------------|----|----|----|----|------|------|------|
| HRV+PTTV     | 34 | 6  | 7  | 33 | 0.85 | 0.83 | 0.84 |
| HRV+PTTV+DDM | 37 | 3  | 5  | 35 | 0.93 | 0.88 | 0.90 |

classifier had a significant increase, with a sensitivity of 0.90, a specificity of 0.85, and an accuracy of 0.88.

#### D. CLASSIFICATION RESULTS FOR SCENARIO III

Table 6 presents the classification results for Scenario III test, where the results were from analysis of HRV, PTTV, as well as the predicted probabilities from DDM-based CNN models. Using only HRV and PTTV indices, SVM classifier generated a sensitivity of 0.85 and a specificity of 0.83, resulting in an accuracy of 0.84. When combined with the predicted probabilities generated from DDM-based CNN models (three for RR and three for PTT time series), the final classification achieved a sensitivity of 0.93 and a specificity of 0.88, giving a highest classification accuracy of 0.90.

#### IV. DISCUSSION

This study combined the information of cardiovascular time series variability and phase space reconstruction to classify normal subjects and heart failure patients. Two time series, i.e., RR and PTT time series were used. HRV and PTTV indices were calculated from the time-, frequency- and non-linear domains. Phase space reconstruction information was from the analysis of DDM by a CNN model. Finally, HRV, PTTV indices and the predicted probabilities from DDM-based CNN models were combined to perform a binary normal or heart failure classification by SVM classifier with LOO-based validation, outputting a sensitivity of 0.93, a specificity of 0.88, and an accuracy of 0.90.

Over the past years, automatic classification for normal and heart failure has been studied. Isler and Kuntalp [18] proposed a model based on KNN and wavelet entropy measures, and achieved accuracies between 78.31% and 84.34% when using different features to train the model. Jovic and Bogunovic [19] proposed a model based on random forest and combinations of linear and non-linear features of HRV, reporting an accuracy of 73% when only using entropy calculation and further improving to around 84% by using combinations of linear and non-linear HRV features. Pecchia *et al.* [20] designed a classifier based on regression tree and the selected HRV features, achieving a sensitivity of 89.74% and a specificity of 100%. However, these high performances were achieved by analyzing the long 24-h

recordings, which is not the situation of short-term analysis in this study. Wang *et al.* [21] still used a SVM method combined with several HRV features on 24-h Holter recording and achieved an accuracy of 90.95%. Our previous work used a combination of CNN and DDM on 24-h RR time series reported an accuracy around 80% for classifying normal sinus rhythm and congestive heart failure [22]. In the current study, we extended the analysis from only using RR time series to using the combination information of RR and PTT time series. One unique new is that, to use the predicted probabilities from DDM-based CNN models as new added input features for SVM-based normal/heart failure classifier. Since the new method involved more information than previous studies, it can achieve high classification accuracy even in the short-term (5-min) application scenario.

As shown in Tables 2 and 3, the majority of HRV indices showed significant differences between two groups while PTTV indices did not, indicating that the sympathetic/parasympathetic nervous system seems like to play only a minor role in the regulation of the peripheral circulation for heart failure patients. The possible explanation can be that blood vessels are mainly innervated by sympathetic nerve fibers, and the change in peripheral sympathetic nerve is not significant in heart failure subjects. However, the SVM classifier using the combination of HRV and PTTV indices obtained a 3% accuracy increase compared with the results from only using HRV indices (shown in Tables 5 and 6), indicating the potential usefulness of PTTV analysis.

In previous study, significant decrease in PTT has been observed in clinical heart failure [38], [39], indicating the heart failure patients had lower peripheral arterial elasticity. The possible reason can be explained that with the decrease of left ventricular function, the activity of peripheral sympathetic nerve is stimulated, and the tension of the artery is enhanced while its elasticity is decreased, resulting in a shorter PTT. The proof can be also from other studies in [40] and [41], where significant increase in PTT was observed within 24 h after percutaneous coronary intervention (PCI) procedure for coronary artery disease patients, indicating that PCI procedure is helpful for improving arterial elasticity and left ventricular functions.

HRV have been widely explored in variety of clinical application, while the study for PTTV is still few. Clear and quantitative PTTV in heart failure keeps unknown. Physiological time series variability, especially the short-term analysis (~ 5 minutes), was welcomed in clinic since the long-term signal recording is difficulty [42]. Recently developed entropy methods have shown good performance for short-term signal analysis [43], [44]. Entropy results in the current study verified the significant reduction of regularity in RR time series while the reduction of regularity in PTT time series was not significant. However, as an intermediate step of entropy calculation, DDM images hold more 2-D information for the time series changes. Thus, adding the DDM information of PTT time series into SVM classifier has enhanced the classification accuracy, generating a

final highest accuracy of 0.90 for normal and heart failure classification.

It is also worth to note that, the calculation method for PTT could generate potential influence on this study. Actually, we need the interval of the arterial pulse wave to travel from the aortic valve to the peripheral arteries, i.e., vascular transit time (VTT). However, the R-peak referred PTT calculation involves both the ventricular pre-ejection period (PEP) and VTT [45], where PEP is an electromechanical delay and is defined as the timing interval between the onsets of ventricular depolarization and ejection. Moreover, the variation in PTT has been proven to tend to follow closely the variation in PEP [46]. Thus, we regarded this point as a limitation of the current study, although obtaining PTT from the R-wave peak to the foot of pulse signal is a common operation in clinical research. In addition, the number of subjects should increase for further confirming the usefulness of the combination of multiple machine learning methods, i.e., CNN+SVM classifiers used in this study.

## V. CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest to this work.

## ACKNOWLEDGMENT

The authors would like to thank the Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab for their support.

## REFERENCES

- [1] A. Mosterd and A. W. Hoes, "Clinical epidemiology of heart failure," *Heart*, vol. 93, no. 9, pp. 1137–1146, 2007.
- [2] F. D. R. Hobbs, J. Doust, J. Mant, and M. R. Cowie, "Heart failure: Diagnosis of heart failure in primary care," *Heart*, vol. 96, no. 21, pp. 1773–1777, 2010.
- [3] A. Fuat, A. P. S. Hungin, and J. J. Murphy, "Barriers to accurate diagnosis and effective management of heart failure in primary care: Qualitative study," *BMJ*, vol. 326, no. 7382, p. 196, 2003.
- [4] M. Hadase *et al.*, "Very low frequency power of heart rate variability is a powerful predictor of clinical prognosis in patients with congestive heart failure," *Circulation J.*, vol. 68, no. 4, pp. 343–347, 2004.
- [5] M. T. La Rovere *et al.*, "Short-term heart rate variability strongly predicts sudden cardiac death in chronic heart failure patients," *Circulation*, vol. 107, no. 4, pp. 565–570, Feb. 2003.
- [6] C.-S. Poon and C. K. Merrill, "Decrease of cardiac chaos in congestive heart failure," *Nature*, vol. 389, no. 6650, pp. 492–495, 1997.
- [7] 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.
- [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] M. D. Thames, T. Kinugawa, M. L. Smith, and M. E. Dibner-Dunlap, "Abnormalities of baroreflex control in heart failure," *J. Amer. College Cardiol.*, vol. 22, no. 4, pp. 56a–60a, Oct. 1993.
- [10] G. Grassi *et al.*, "Sympathetic activation and loss of reflex sympathetic control in mild congestive heart failure," *Circulation*, vol. 92, no. 11, pp. 3206–3211, Dec. 1995.
- [11] P. Van De Borne, N. Montano, M. Pagani, R. Oren, and V. K. Somers, "Absence of low-frequency variability of sympathetic nerve activity in severe heart failure," *Circulation*, vol. 95, no. 6, pp. 1449–1454, Mar. 1997.
- [12] J. Nolan *et al.*, "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 (UK-heart)," *Circulat.*, vol. 98, no. 15, pp. 1510–1516, 1998.
- [13] R. Maestri *et al.*, "Nonlinear indices of heart rate variability in chronic heart failure patients: Redundancy and comparative clinical value," *J. Cardiovascular Electrophysiol.*, vol. 18, no. 4, pp. 425–433, 2000.
- [14] T. H. Mäkikallio *et al.*, "Fractal analysis and time- and frequency-domain measures of heart rate variability as predictors of mortality in patients with heart failure," *Amer. J. Cardiol.*, vol. 87, no. 2, pp. 178–182, 2001.
- [15] 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.
- [16] C. Y. Liu *et al.*, "Comparison of different threshold values  $r$  for approximate entropy: Application to investigate the heart rate variability between heart failure and healthy control groups," *Physiol. Meas.*, vol. 32, no. 2, pp. 167–180, 2011.
- [17] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, p. 068102, Jul. 2002.
- [18] Y. İşler and M. Kuntalp, "Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure," *Comput. Biol. Med.*, vol. 37, no. 10, pp. 1502–1510, 2007.
- [19] A. Jovic and N. Bogunovic, "Random forest-based classification of heart rate variability signals by using combinations of linear and nonlinear features," in *Proc. 12th Medit. Conf. Med. Biol. Eng. Comput.* Berlin, Germany: Springer, 2010, pp. 29–32.
- [20] L. Pecchia, P. Melillo, M. Sansone, and M. Bracale, "Discrimination power of short-term heart rate variability measures for CHF assessment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 40–46, Jan. 2010.
- [21] Y. Wang *et al.*, "Comparison of time-domain, frequency-domain and nonlinear analysis for distinguishing congestive heart failure patients from normal sinus rhythm subjects," *Biomed. Signal Process. Control*, vol. 42, no. 4, pp. 30–36, 2018.
- [22] Y. W. Li *et al.*, "Combining convolutional neural network and distance distribution matrix for identification of congestive heart failure," *IEEE Access*, vol. 6, pp. 39734–39744, 2018.
- [23] A. Bravi, A. Longtin, and A. J. Seely, "Review and classification of variability analysis techniques with clinical applications," *Biomed. Eng. Online*, vol. 10, no. 10, p. 90, 2011.
- [24] C. Liu, C. Zhang, L. Zhang, L. Zhao, C. Liu, and H. Wang, "Measuring synchronization in coupled simulation and coupled cardiovascular time series: A comparison of different cross entropy measures," *Biomed. Signal Process. Control*, vol. 21, no. 8, pp. 49–57, 2015.
- [25] S. Torraca *et al.*, "Variability of pulse wave velocity and mortality in chronic hemodialysis patients," *Hemodialysis Int.*, vol. 15, no. 3, pp. 326–333, 2011.
- [26] C. H. Tang *et al.*, "Pulse transit time variability analysis in an animal model of endotoxic shock," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2010, pp. 2849–2852.
- [27] R. V. Carlson, K. M. Boyd, and D. J. Webb, "The revision of the Declaration of Helsinki: Past, present and future," *Brit. J. Clin. Pharmacol.*, vol. 57, no. 6, pp. 695–713, 2004.
- [28] C. Liu *et al.*, "Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2844090.
- [29] C. Liu, Q. Li, and G. D. Clifford, "Evaluation of the accuracy and noise response of an open-source pulse onset detection algorithm on pulsatile waveform databases," in *Proc. Comput. Cardiol.*, Vancouver, BC, Canada, 2016, pp. 913–916.
- [30] M. P. Tulppo, T. H. Mäkikallio, T. E. Takala, T. Seppänen, and H. V. Huikuri, "Quantitative beat-to-beat analysis of heart rate dynamics during exercise," *Amer. J. Physiol.*, vol. 271, no. 1, pp. H244–H252, Jul. 1996.
- [31] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiology: Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [32] C. Liu *et al.*, "Analysis of heart rate variability using fuzzy measure entropy," *Comput. Biol. Med.*, vol. 43, no. 2, pp. 100–108, 2013.
- [33] C. Liu and L. Zhao, "Using fuzzy measure entropy to improve the stability of traditional entropy measures," in *Proc. Comput. Cardiol.*, Hangzhou, China, 2011, pp. 681–684.



- [34] C. Liu and A. Murray, "Applications of complexity analysis in clinical heart failure," in *Complexity and Nonlinearity in Cardiovascular Signals*, R. Barbieri, E. Scilingo, and G. Valenza, Eds. Cham, Switzerland: Springer, 2017, pp. 301–325.
- [35] L. Zhao et al., "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.
- [36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [37] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27–1–27–27, 2011.
- [38] C. Liu et al., "Elastic properties of peripheral arteries in heart failure patients in comparison with normal subjects," *J. Physiol. Sci.*, vol. 63, no. 3, pp. 195–201, 2013.
- [39] C. Giannattasio et al., "Radial, carotid and aortic distensibility in congestive heart failure: Effects of high-dose angiotensin-converting enzyme inhibitor or low-dose association with angiotensin type 1 receptor blockade," *J. Amer. College Cardiol.*, vol. 39, pp. 1275–1282, 2002.
- [40] G. Zhang, C. Liu, L. Ji, J. Yang, and C. Liu, "Effect of a percutaneous coronary intervention procedure on heart rate variability and pulse transit time variability: A comparison study based on fuzzy measure entropy," *Entropy*, vol. 18, no. 7, p. 246, 2016.
- [41] L. Ji, C. Liu, X. Wang, Y. Hou, and C. Liu, "Increased pulse wave transit time after percutaneous coronary intervention procedure in CAD patients," *Sci. Rep.*, vol. 8, Jan. 2018, Art. no. 115.
- [42] J. T. Bigger, J. L. Fleiss, L. M. Rolnitzky, and R. C. Steinman, "The ability of several short-term measures of RR variability to predict mortality after myocardial infarction," *Circulation*, vol. 88, no. 3, pp. 927–934, 1993.
- [43] C. Liu et al., "A comparison of entropy approaches for AF discrimination," *Physiol. Meas.*, vol. 39, no. 7, pp. 74002–1–74002–18, 2018.
- [44] A. Humeau-Heurtier, "The multiscale entropy algorithm and its variants: A review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, 2015.
- [45] Q. Li and G. G. Belz, "Systolic time intervals in clinical pharmacology," *Eur. J. Clin. Pharmacol.*, vol. 44, no. 5, pp. 415–421, 1993.
- [46] G. S. Chan, P. M. Middleton, B. G. Celler, L. Wang, and N. H. Lovell, "Change in pulse transit time and pre-ejection period during head-up tilt-induced progressive central hypovolaemia," *J. Clin. Monit. Comput.*, vol. 21, no. 5, pp. 283–293, Oct. 2007.



**LINA ZHAO** received the B.S. and M.S. degrees in biomedical engineering from Shandong University, Jinan, China, in 2005 and 2008, respectively, where she is currently pursuing the Ph.D. degree with the School of Control Science and Engineering. She was as a Researcher with Shandong Heng-Xin Inspection Technique Exploiture Center, Jinan. Her research interests include entropy methods for physiological signal analysis, ECG, PCG, and artery pressure pulse processing.



**CHENGYU LIU** (M'14) received the B.S. and Ph.D. degrees in biomedical engineering from Shandong University, China, in 2005 and 2010, respectively. He has completed the Postdoctoral trainings at Shandong University, from 2010 to 2013, Newcastle University, U.K., from 2013 to 2014, and Emory University, USA, from 2015 to 2017. He was the PI on over ten awarded grants attracting a total of over \$1 million. He is currently the Director and a Professor with Southeast-

Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab, School of Instrument Science and Engineering, Southeast University, Nanjing, China. He has published more than 150 journal/conference papers and eight chapters in books. He holds 15 invention patents. His research interests include wearable ECG monitoring, machine learning and big data processing for physiological signals, early detection of CADs, device development for CADs, and sleep and emotion monitoring. He is a Federation Journal Committee Member of the International Federation for Medical and Biological Engineering.



**SHOUSHUI WEI** received the M.S. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1992. He is currently a Professor with the Institute of Biomedical Engineering, Shandong University, China. His research interests include ECG signal processing, machine learning for physiological signals analysis, and early detection of CADs.



**CHANGCHUN LIU** received the B.S. and M.S. degrees in automatic control from Shandong University, Jinan, China, in 1982 and 1987, respectively, where he is currently a Professor of biomedical engineering. His research interests include biomedical signal and image processing, biomedical measurements and devices, and physiological system modeling.



**JIANQING LI** received the B.S. and M.S. degrees in automatic technology from the School of Instrument Science and Engineering, Southeast University, Nanjing, China, in 1986 and 1992, respectively, and the Ph.D. degree in measurement technology and instruments from Southeast University. He has been with Southeast University, since 1986, where he is currently a Professor with the School of Instrument Science and Engineering. He is also a Professor with the School of Basic Medical Sciences, Nanjing Medical University. His research interests include wearable ECG processing and smart health.

...