Received: 25 November 2018 | Revised: 31 January 2019 | Accepted: 22 February 2019

DOI: 10.1111/exsy.12396

Check for updates

ORIGINAL ARTICLE

WILEY Expert Systems

Classification of congestive heart failure with different New York Heart Association functional classes based on heart rate variability indices and machine learning

Zhaohui Qu† | Qianwen Liu† | Chengyu Liu

School of Instrument Science and Engineering, Southeast University, Nanjing, China

Correspondence
Chengyu Liu, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China.
Email: chengyu@seu.edu.cn

Funding information
Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab; Primary Research & Development Plan of Jiangsu Province, Grant/Award Number: BE2017735; National Natural Science Foundation of China, Grant/Award Number: 81871444

Abstract

This study aims to evaluate the effect of heart rate variability (HRV) indices on the New York Heart Association (NYHA) classification of patients with congestive heart failure and to test the effectiveness of different machine learning algorithms. Twenty-nine long-term RR interval recordings from subjects (aged 34 to 79) with congestive heart failure (NYHA classes I, II, and III) in MIT-BIH Database were studied. We firstly removed the unreasonable RR intervals and segment the RR recordings with a 300-RR interval length window. Then the multiple HRV indexes were calculated for each RR segment. Support vector machine (SVM) and classification and regression tree (CART) methods were then separately used to distinguish patients with different NYHA classes based on the selected HRV indices. Receiver operating characteristic curve analysis was finally employed as the evaluation indicator to compare the performance of the two classifiers. The SVM classifier achieved accuracy, sensitivity, and specificity of 84.0%, 71.2%, and 83.4%, respectively, whereas the CART classifier achieved 81.4%, 66.5%, and 81.6%, respectively. The area under the curve of receiver operating characteristic for the two classifiers was 86.4% and 84.7%, respectively. It is possible for accurately classifying the NYHA functional classes I, II, and III when using the combination of HRV indices and machine learning algorithms. The SVM classifier performed better in classification than the CART classifier using the same HRV indices.

KEYWORDS

cardiovascular disease, classification, congestive heart failure, electrocardiogram, heart rate variability

1 | INTRODUCTION

Recent studies indicated that heart rate variability (HRV) plays an important role in diagnosis and treatment in cardiovascular disease such as congestive heart failure (CHF; Malik, 1998). CHF is a disorder in which the heart cannot pump blood efficiently thus causing many organs to lack of enough oxygen and nutrients (İşler & Kuntalp, 2007). CHF is a difficult condition in clinical treatment and has a high mortality rate (Cohn, 1996; Nolan et al., 1998). The New York Heart Association (NYHA) introduced the NYHA functional classification system to help physicians in clinical practice evaluate the effect of cardiac symptoms on a patient's daily activities (Bennett, Riegel, Bittner, & Nichols, 2002).

†These authors contributed equally in this study.

 © 2019 John Wiley & Sons, Ltd

HRV analysis is of great significance for the diagnosis of CHF (Sztajzel, 2004). Ponikowski et al. (1997) studied 102 consecutive patients with moderate to severe CHF and concluded that depressed HRV could be an independent predictor for a poor prognosis in patients with CHF (Ponikowski et al., 1997). Stein, Rich, Rottman, and Kleiger (1995) measured Holter-derived HRV indexes in 17 stable patients with class II or III CHF and concluded that time- and frequency-domain indexes of HRV showed stability in CHF patients over time (Stein et al., 1995). Nolan et al. (1998) recruited 433 outpatients with CHF (NYHA functional classes I to III) and found that the standard deviation (SDNN) of RR intervals reduced in patients at high risk of death (Nolan et al., 1998). Hadase et al. (2004) monitored 54 consecutive CHF patients and reported that very low-frequency power is an independent risk predictor of clinical prognosis in patients with CHF (Hadase et al., 2004). Ho et al. (1997) studied HRV in 2-hr ambulatory electrocardiogram (ECG) recordings of 69 participants and put forward that non-linear HRV indices may provide complementary help to traditional HRV measures in treating CHF patients (Ho et al., 1997). Costa, Goldberger, and Peng (2002) proved that multiscale entropy could separate healthy and CHF groups robustly against the influence of noise (Costa et al., 2002). Signorini, Ferrario, Marchetti, and Marseglia (2006) successfully applied approximate entropy (ApEn) and sample entropy (SampEn) to differentiate between CHF and normal subjects (Signorini et al., 2006). Graff, Graff, and Kaczkowska (2012) computed ApEn, SampEn, fuzzy entropy, and permutation entropy on RR intervals of different lengths, finding that ApEn, SampEn, fuzzy entropy, and permutation entropy were significantly different between patients with CHF and healthy individuals (Graff et al., 2012). D'addio et al. (1998) performed frequency-domain and a Poincare plot analysis on 22 CHF patients and confirmed that they might be useful in clinical setting (D'addio et al., 1998). Kamen and Tonkin (1995) proposed a morphological classification scheme using Poincare plot pattern to display HRV data from a group of 23 patients with heart failure and found a significant difference between patients with NYHA classes I and II compared with patients with NYHA classes III and IV (Kamen & Tonkin, 1995).

At present, machine learning method can reorganize existing knowledge structures by using inductive, integrated rather than deductive methods to acquire new knowledge or skills and continually improve their performance (Alpaydin, 2009). Supervised learning, an important branch of machine learning, can be used in the diagnosis and prediction of cardiovascular diseases in clinical work. It has become a common tool in detecting CHF. Pecchia, Melillo, and Bracale (2011) proposed a platform based on the classification and regression tree (CART) method and reported a good performance in detecting heart failure (Pecchia, Melillo, & Bracale, 2011). Pecchia, Melillo, Sansone, and Bracale (2011) studied 83 patients to investigate the discrimination power of short-term HRV for discriminating normal subjects versus CHF patients. The CART classifier they used achieved high sensitivity and specificity (Pecchia, Melillo, Sansone, & Bracale, 2011). Khaled, Owis, and Mohamed (2006) employed time-domain methods and Poincare plot to detect CHF using minimum-distance classifier, Bayes minimum-error classifier, voting k-nearest neighbour classifier, and Back propagation neural networks. Time-domain indices were proved to be more capable of discriminating the normal from CHF signals than Poincare plot features (Khaled et al., 2006). Mehrabi, Maghsoudloo, Arabalibeik, Noormand, and Nozari (2009) differentiated between patients suffering from CHF and chronic obstructive pulmonary disease. They employed multilayer perceptron and radial basis function neural networks and finally obtained fine classification (Mehrabi et al., 2009). Masetic and Subasi (2016) compared five different classifiers (C4.5 decision tree, k-nearest neighbour, support vector machine [SVM], artificial neural networks, and random forest [RF] classifier) in detecting CHF. They performed their experiment on the ECG signals acquired from BIDMC CHF and PTB diagnostic ECG databases and found that RF method was the most powerful classifier (Masetic & Subasi, 2016).

Previous studies have focused primarily on discriminating CHF from normal cardiac function, whereas detailed analysis of the effect of HRV indices on distinguishing the different degrees of CHF is lacking. The main goal of this study is to evaluate the effect of HRV indices on the NYHA classification of patients with CHF. In this study, both the SVM and CART classifiers were used to test the classification effect of time-domain, frequency-domain, and non-linear HRV indices for distinguishing different NYHA levels of CHF (NYHA classes I, II, and III). Receiver operating characteristic (ROC) curve analysis was introduced to compare the performance of the two classifiers. The remainder of this paper was organized as follows. Our method, including data preprocessing and algorithms employed to implement a classifier, was presented in Section 2. Corresponding results were shown in Section 3. We gave a discussion in Section 4.

2 | METHODS

2.1 | Data

All data used here were from the RR Interval Databases from http://www.physionet.org (Goldberger et al., 2000). The RR Interval Database included 29 long-term RR interval recordings of subjects (aged 34 to 79) with CHF, 4 labelled as NYHA functional class I, 8 labelled as class II, and 17 labelled as class III. The sample frequency of the raw ECG signal is 128 Hz. We used the QRS annotations to obtain the RR interval time series, and these annotations were from automated analysis with manual review and correction.

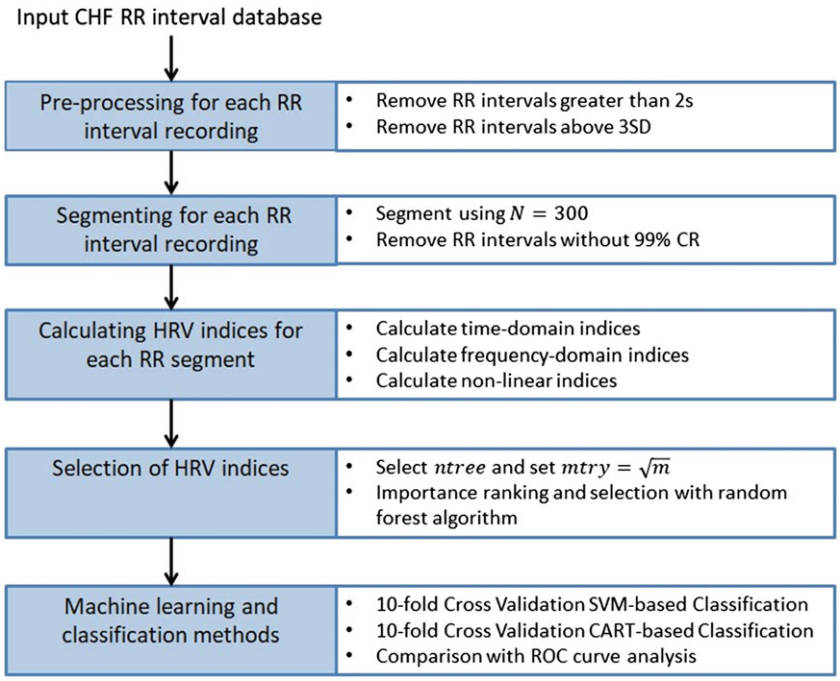

FIGURE 1 Block diagram of the proposed analytical procedure. CART: classification and regression tree; CHF: congestive heart failure; HRV: heart rate variability; ROC: receiver operating characteristic; SVM: support vector machine

2.2 | Method description

Figure 1 shows the block diagram of the analytical procedure used in this paper. There are five major steps in this procedure: Step 1, preprocessing for each RR interval recording; Step 2, segmenting for each RR interval recording; Step 3, calculating HRV indices for each RR segment; Step 4, selection of HRV indices; and Step 5, machine learning and classification methods.

In Step 1, to avoid the deviations brought by artefacts, the RR intervals greater than 2 s are removed from original RR intervals (C. Liu & Gao, 2017). Then the RR intervals that present above three standard deviations (3SD) around the mean are identified and removed (Berntson & Stowell, 1998). Table 1 shows the number of RR intervals in original CHF groups (NYHA classes I, II, and III) and the numbers of RR intervals after the preprocessing operations mentioned above. The percentages of the reserved RR intervals after the removal are 99.3%, 97.1%, and 97.5% for NYHA classes I, II, and III groups, respectively.

In Step 2, a 300-RR interval window is utilized to segment the long-term RR intervals and form the RR segments. Then we refuse RR intervals whose confidence intervals are below 99% for each RR segment to further reduce interferes brought by artefacts and ectopic beats (Christou & Dinov, 2011).

In Step 3, we calculated the common HRV indices for each RR segment, including time-domain indices, frequency-domain indices, and nonlinear indices. The detailed descriptions of these indices are shown in Section 2.3.

In Step 4, we use RF algorithm against overfitting to perform importance ranking and selection of HRV indices. The final indices we chose were the subset of features corresponding to the smallest out-of-bag (OOB) error (Breiman, 2001; Genuer, Poggi, & Tuleau-Malot, 2010). We set the

TABLE 1 Statistical profile for the 29 CHF RR Interval Databases

	NYHA functional classification		
Variable	**Class I**	**Class II**	**Class III**
NRR interval recordings	chf213, chf217 chf218, chf221	chf211, chf212, chf214 chf215, chf216, chf220 chf224, chf226	chf201–chf210, chf219 chf222, chf223, chf225 chf227, chf228, chf229
No. of RR interval recordings	4	8	17
No. of RR intervals	428,723	955,867	1,905,752
No. of RR intervals after removing greater than 2 s	427,761	954,657	1,902,029
No. of RR intervals after removing abnormal heartbeats	425,668	927,798	1,857,333
No. of RR segments when setting N = 300	1,416	3,088	6,181

Note. CHF: congestive heart failure; NYHA: New York Heart Association.

number of random variables dividing the node *mtry* = $\sqrt{m}$, where *m* is the number of HRV indices of all the segments, according to Liaw and Wiener (2002). The number of trees in the forest *ntree* is the one that yields to the smallest error rate. That is to say, the value of *ntree* we decided should guarantee the convergence of RF algorithm (Archer & Kimes, 2008). The selected HRV indices will be used as input for classification.

In Step 5, we use the statistical methods shown in Section 2.5 to compare the HRV results among the NYHA classes I, II, and III groups. Then we used an SVM-based classifier and a CART-based classifier separately to classify the NYHA classes I, II, and III groups. ROC curve analysis was used for algorithm evaluation. The detailed descriptions of machine learning and classification methods were summarized in Section 2.4.

2.3 | HRV indices

2.3.1 | Time-domain indices

The mean value (MEAN) of RR intervals, the SDNN of RR intervals, the square root of mean squared differences of successive RR intervals (RMSSD), and the proportion derived by dividing NN50 (the number of interval differences of successive NN intervals greater than 50 ms) by the total number of NN intervals (PNN50) were used as time-domain indices (Brennan, Palaniswami, & Kamen, 2001; Camm et al., 1996). SDNN and RMSSD measured the overall regulation of the autonomic nervous system and vagus nerve on heart rate, respectively. PNN50 showed the tension of the vagus nerve.

2.3.2 | Frequency-domain indices

The Burg's method is employed to get the parameter of the AR model, which is used for the analysis of frequency domain (Acharya, Joseph, Kannathal, Min, & Suri, 2007). Using Burg's method with an order of 16, we could produce the spectrum of HRV frequency spectrum that was integrated across the low-frequency power (0.04 to 0.15 Hz) and high-frequency power (0.15 to 0.40 Hz) spectra (Brennan et al., 2001; Camm et al., 1996). The representation of LF (low-frequency power) and HF (high-frequency power) in normalized units reflects the controlled behaviour of sympathetic and parasympathetic, respectively, whereas their ratio (LF/HF) emphasized the balanced behaviour of the two branches of the autonomic nervous system (Malik et al., 1996). Here, LFn = LF/(HF + LF), HFn = HF/(HF + LF), and their ratio (LF/HF) were used as the frequency-domain indices.

2.3.3 | Non-linear indices

We introduce SampEn and fuzzy measure entropy (FuzzyMEn) as non-linear indices. The value of SampEn reflects the complexity of the sequence (Richman & Moorman, 2000). Due to the improved accuracy of SampEn statistics, they can be used in the study of experimental clinical cardiovascular and other biological time series. Combining both local similarity and global similarity in time series, FuzzyMEn also has a good discrimination for time series with inherent complexity. Following the parameter settings in Zhao et al. (2015), the settings for SampEn were embedding dimension $m = 2$ and tolerance threshold $r = 0.1$; the parameter setting for FuzzyMEn is embedding dimension $m = 2$.

Figure 2 shows the example outputs of the aforementioned HRV indices. Similar results among the three groups could be observed.

2.4 | Machine learning and classification methods

2.4.1 | SVM-based classifier

For binary classification problem, the standard SVM can be used (Blanz et al., 1996). For multi-class classification problem, one method is to decompose standard SVMs to several binary classification problems; therefore, a standard SVM can be used (Franc & Hlavác, 2002). Here, one-against-all decomposition is applied based on the libsvm software package to learn the SVM models (Chang & Lin, 2011). A 10-fold crossvalidation is performed to the CHF subjects, that is, RR segments of all the patients' recordings were randomly divided to 10 folds. We check onefold at a time. All subjects that are not in the current fold formed the training set to train the SVM. The test set formed by subjects in the current fold is used to test the result of NYHA classes I, II, and III classification. It is a total patient-independent test. The parameter settings of libsvm are listed as follows. Radial basis function is used as kernel function according to Keerthi and Lin (2003) and Lin and Lin (2003). A "grid-search" on cost parameter *C* and gamma parameter γ using crossvalidation is used (Hsu, Chang, & Lin, 2003). We tried various pairs of (C, γ) values and found that the best accuracy reached when $\gamma = 0.000122703125$ and $C = 2$. Sensitivity (*Se*), specificity (*Sp*), and accuracy (*Acc*) were calculated to evaluate the SVM model. Sensitivity, specificity, and accuracy were defined as

$$Se = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}), \tag{1}$$

$$Sp = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP}), \tag{2}$$

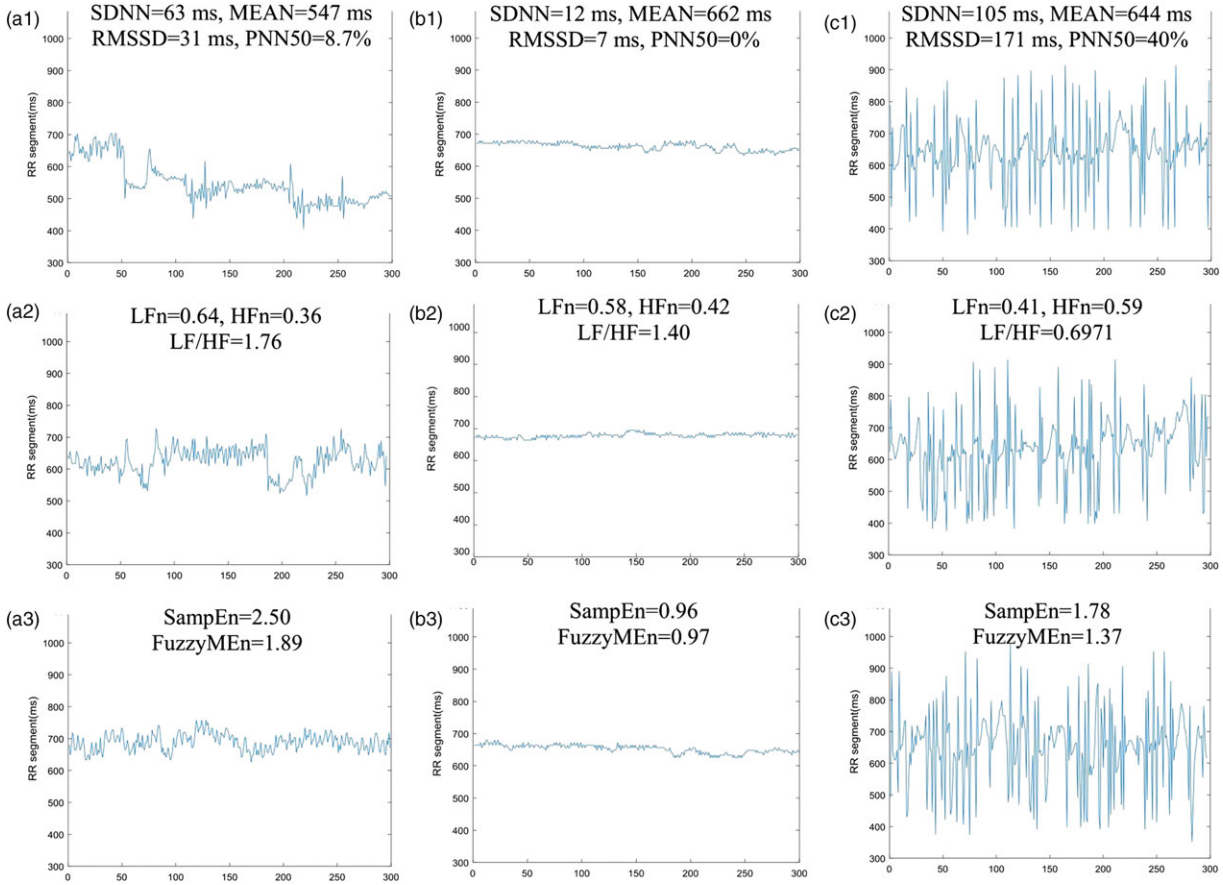

FIGURE 2 Examples of HRV results from NYHA functional class III patient (a1–a3), class II patient (b1–b3), and class I patient (c1–c3). The figures (a1–a3) were the 29th, the 19th, and the 6th segment from the recordings of subject chf29, respectively. The waveforms (b1–b3) were the 4th, the 11th, and the 14th segment from the patient chf26's recordings, respectively. The waveforms (c1–c3) were the 13th, the 14th, and the 15th segment from the subject chf26's recordings, separately

$$Acc = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}), \tag{3}$$

where for NYHA functional class I group, TP is the number of the patients with class I that were correctly classified as the class I group, TN is the number of other patients that were classified as class II or III groups, FP is the number of other patients that were falsely classified as the class I group, and FN is the number of the patients with class I that were falsely classified as class II or III groups. For class II and III groups, the meanings of these notations can be deduced by analogy.

2.4.2 | CART-based classifier

CART is a classification method that uses historical statistics to construct decision trees. The segregation of different values of classification variables through a decision tree composed of progressive binary splits is often involved in such algorithm (Fonarow et al., 2005). With HRV indices values as the sample dataset and a 10-fold validation performed to the subjects, the training process here is the same as Section 2.4.1.

Gini index was used as the criterion for selecting the best splitting attribute. For any subset $\boldsymbol{Sq}$ of training dataset, it is given by Rutkowski, Jaworski, Pietruczuk, and Duda (2014).

$$\mathrm{Gini}(\boldsymbol{Sq}) = 1 - \sum_{k=1}^{K} p_{k,q}{}^2, \tag{4}$$

where K is the number of category (here, K = 3) and $p_{k,q}$ is the probability that category k appears in $\boldsymbol{Sq}$. The Gini index is calculated separately for the combination of the binary divisions, and the combination of the minimized Gini index is automatically selected. To evaluate performance, sensitivity, specificity, and accuracy, as defined in Equations 1–3, respectively, were still used.

2.4.3 | ROC curve analysis

ROC is one standard tool that is used to analyse and compare classifiers in the condition of the datasets with unbalanced classes or unknown cost of misclassification (Clearwater & Stern, 1991). In this paper, we employ the common index of area under the curve (AUC) to evaluate the classification performances of SVM-based and CART-based algorithms and determine the better choice for distinguishing the CHF patients of different NYHA classes.

2.5 | Statistical analysis

We employ Kolmogorov–Smirnov test and Q–Q plot to test the normal distributions of all HRV indices. If passed, the group *t* test was used to test the statistical difference among the NYHA classes I, II, and III groups. Otherwise, non-parametric test was used instead. All statistical results were considered statistically significant if *p* values are less than 0.05. We perform all our statistical analysis on the SPSS software (Ver. 19, IBM, USA).

3 | RESULTS

3.1 | Statistical differences of HRV indices among the three groups

We first calculated the HRV indices for all the segments. All HRV indices of CHF patients contended normal distribution according to the result of the Kolmogorov–Smirnov test. Then we got the mean ± standard error of the mean values for NYHA functional classes I, II, and III groups and the whole CHF patients group. Table 2 shows the mean ± standard error of the mean values and the lower and upper bounds of normalized 95% confidence interval of mean group values for the three groups separately and the whole CHF subjects. Table 3 shows the statistical results for the three groups.

3.2 | Selection of HRV indices

3.2.1 | Selection of *ntree*

We gradually increased the number of trees in forest from 0 to 1,000 and calculated the OOB error of all HRV indicators. Figure 3 illustrated the change of OOB error of RF model with *ntree* growing. As is shown in the figure, OOB error keeps stable at 23% when *ntree* is equal to 200 or more. Thus, *ntree* = 200 was chosen for the RF method.

3.2.2 | Importance ranking and selection of HRV indices

With RF algorithm, we got the ranking list of the nine HRV indices and sorted them from high to low according to the importance level. Figure 4 shows the importance ranking of these indices. The importance level values of MEAN, RMSSD, SDNN, PNN50, LF/(LF + HF), HF/(LF + HF), LF/HF, SampEn, and FuzzyMEn are 0.1950, 0.1593, 0.1316, 0.0849, 0.0693, 0.0697, 0.0712, 0.1196, and 0.0994, respectively. The subset of features corresponding to the smallest OOB error was made up of all the time-domain and non-linear indices, as well as the frequency-domain LF/HF. Therefore, the HRV indices LF/(LF + HF) and HF/(LF + HF) were refused in subsequent studies.

3.3 | Classification results using SVM classifier and CART classifier

Table 4 shows the classification results of 29 CHF subjects with three different NYHA classes using 10-fold crossvalidation-based SVM classifier. The average *Se*, *Sp*, and *Acc* of class I are 69.6%, 97.4%, and 92.9%, respectively; those of class II are 52.6%, 93.2%, and 80.9%, respectively; and those of class III are 91.3%, 59.7%, and 78.1%, respectively, resulting in an average *Acc* of 84.0% for all classes.

Table 5 shows the classification results of 29 CHF subjects with three different NYHA classes using 10-fold crossvalidation-based CART classifier. The average *Se*, *Sp*, and *Acc* of class I are 56.5%, 96.7%, and 90.2%, respectively; those of class II are 55.4%, 89.8%, and 78.5%, respectively; and those of class III are 87.6%, 58.2%, and 75.5%, respectively, resulting in an average *Acc* of 81.4% for all classes.

3.4 | Comparison of the two classifiers with ROC curve analysis

Figure 5 shows ROC curve plots with AUC values for the results of 10-fold crossvalidation-based SVM and CART classifiers for classifying NYHA functional classes I, II, and III groups. The SVM classifier achieved AUC values of 95.2%, 84.5%, and 79.5% for classifying classes I, II, and III, respectively, whereas the AUC values of CART classifier were 89.4%, 79.4%, and 85.2%, respectively. It is clear that the area under the ROC curve of the SVM classifier is larger than that of CART classifier in each figure.

4 | DISCUSSIONS

Decreased HRV indices were regarded as independent risk factors for mortality in patients with advanced CHF (Aronson, Mittleman, & Burger, 2004; Ponikowski et al., 1997; Stein, Bosner, Kleiger, & Conger, 1994). Lee, Noh, and Ryu (2007) evaluated several classifiers with various

TABLE 2 Statistical results of HRV indices for the three NYHA functional class groups

NYHA class	HRV indices	Mean ± SEM	95% CI of mean group values	
			Lower	Upper
Class I	MEAN (ms)	762.05 ± 4.00	754.36	769.57
	SDNN (ms)	76.09 ± 1.80	72.71	79.71
	RMSSD (ms)	70.76 ± 1.32	68.30	73.50
	PNN50	16.90 ± 0.44	16.03	17.75
	LF/LH	1.64 ± 0.05	1.55	1.75
	LFn	0.49 ± 0.01	0.48	0.50
	HFn	0.51 ± 0.01	0.50	0.52
	SampEN	1.63 ± 0.01	1.61	1.66
	FuzzyMEn	0.97 ± 0.01	0.95	1.00
Class II	MEAN (ms)	637.97 ± 1.59	634.70	641.15
	SDNN (ms)	18.82 ± 0.26	18.32	19.34
	RMSSD (ms)	24.25 ± 0.31	23.65	24.82
	PNN50	3.53 ± 0.14	3.27	3.80
	LF/LH	1.82 ± 0.03	1.75	1.88
	LFn	0.52 ± 0.00	0.52	0.53
	HFn	0.48 ± 0.00	0.47	0.48
	SampEN	1.44 ± 0.01	1.42	1.45
	FuzzyMEn	1.23 ± 0.01	1.21	1.25
Class III	MEAN (ms)	687.92 ± 1.31	685.47	690.54
	SDNN (ms)	24.41 ± 0.36	23.69	25.14
	RMSSD (ms)	26.97 ± 0.26	26.45	27.47
	PNN50	4.76 ± 0.13	4.50	5.02
	LF/LH	1.20 ± 0.02	1.16	1.23
	LFn	0.45 ± 0.00	0.44	0.45
	HFn	0.55 ± 0.00	0.55	0.56
	SampEN	1.52 ± 0.01	1.51	1.53
	FuzzyMEn	1.33 ± 0.01	1.31	1.34
All	MEAN (ms)	683.30 ± 1.08	681.34	685.40
	SDNN (ms)	29.64 ± 0.36	28.95	30.38
	RMSSD (ms)	31.98 ± 0.28	31.44	32.53
	PNN50	6.01 ± 0.11	5.79	6.23
	LF/LH	1.43 ± 0.02	1.40	1.46
	LFn	0.47 ± 0.00	0.47	0.48
	HFn	0.53 ± 0.00	0.52	0.53
	SampEN	1.51 ± 0.00	1.50	1.52
	FuzzyMEn	1.25 ± 0.01	1.24	1.26

Note. CI: confidence interval; HRV: heart rate variability; NYHA: New York Heart Association; SEM: standard error of the mean.

experiments on linear and non-linear features of HRV indices and found that SVM performed better than other classifiers such as CART and C4.5 (Lee et al., 2007). İşler and Kuntalp (2007) performed a k-nearest-neighbour classifier to evaluate the performance of feature sets of HRV indices in distinguish CHF group from the normal subjects and reported that the best set was made up of MEAN, RMSSD, VLI, LF/HF, very low frequency (0–0.04 Hz), LFn, and HFn. The classifier obtained the sensitivity and specificity value of 100% and 94.74% (İşler and Kuntalp, 2007). Pecchia, Melillo, and Bracale (2011) distinguished normal subjects versus CHF patients with CART method based on short-term HRV indices, included RMSSD, total power, HFn, and LF/HF. The classifier obtained sensitivity and specificity values of 79.3% and 100%, respectively (Pecchia, Melillo, Sansone, & Bracale, 2011). Melillo, De Luca, Bracale, and Pecchia (2013) achieved a sensitivity and a specificity rate of 93.3% and 63.6%, respectively, in identifying higher risk patients with CHF (NYHA functional classes III and IV) using CART method based on standard long-term HRV measures (Melillo et al., 2013). G. Liu et al. (2014) achieved the best performance with the CHF classification accuracy, sensitivity, and specificity of 100%, 100%, and 100%, respectively, based on SVM algorithm and two HRV indices SUM_TD and SUM_FD (G. Liu et al., 2014). However, these results are not patient independent, with the risk of overfitting.

NYHA functional classification has been proved to be a risk factor for mortality in CHF patients (Senni et al., 1998). All HRV indices used in this study showed statistical differences between any two of the three NYHA functional class groups. The difference among the three groups could be explained by the decreased parasympathetic activity to the heart and the vagal outflow of the heart (Casolo, Balli, Taddei, Amuhasi, & Gori, 1989). Both the SVM-based and CART-based classifiers obtained high accuracy and performed well in ROC curve analysis when classifying the three groups. The average value of *Acc* of the two classifiers was 84.0% and 81.4%, respectively. The 10-fold crossvalidation-based SVM and CART obtained average AUC of 86.4% and 84.7% separately. However, the sensitivity of the two classifiers was not that high when detecting classes I and II (<60%) due to the slight differences in clinical practice of these two classes. However, SVM classifier performs more powerful in discriminating classes I, II, and III groups than CART classifier when using the same HRV indices.

In short, this study shows that it is possible for accurately classifying the NYHA functional classes I, II, and III when using the combination of HRV indices and machine learning algorithms. The SVM classifier performed better in classification than the CART classifier using the same HRV indices.

TABLE 3 Statistical significances of any two of the three New York Heart Association functional class groups

HRV indices	(A) class	(B) class	*p* value
MEAN (ms)	Class I	Class II	0.00
	Class I	Class III	0.00
	Class II	Class III	0.00
RMSSD (ms)	Class I	Class II	0.00
	Class I	Class III	0.00
	Class II	Class III	0.00
SDNN (ms)	Class I	Class II	0.00
	Class I	Class III	0.00
	Class II	Class III	2.79×10^{-11}
PNN50	Class I	Class II	0.00
	Class I	Class III	0.00
	Class II	Class III	1.40×10^{-10}
LF/HF	Class I	Class II	0.0146
	Class I	Class III	0.00
	Class II	Class III	0.00
LFn	Class I	Class II	1.17×10^{-7}
	Class I	Class III	2.15×10^{-9}
	Class II	Class III	0.00
HFn	Class I	Class II	1.17×10^{-7}
	Class I	Class III	2.15×10^{-9}
	Class II	Class III	0.00
SampEn	Class I	Class II	0.00
	Class I	Class III	6.66×10^{-16}
	Class II	Class III	0.00
FuzzyMEn	Class I	Class II	0.00
	Class I	Class III	0.00
	Class II	Class III	2.00×10^{-15}

Note. HRV: heart rate variability.

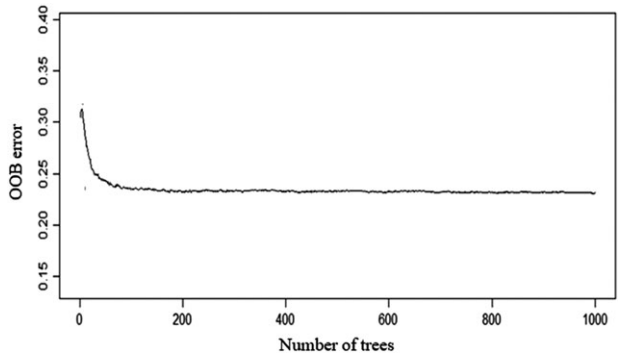

FIGURE 3 Change of out-of-bag (OOB) error of random forest model with *ntree* growing from 0 to 1,000. The value of OOB error declines as *ntree* increases

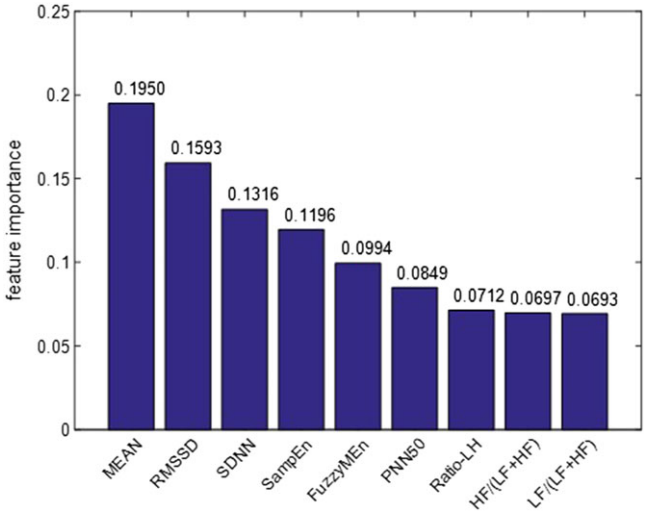

FIGURE 4 Importance ranking of heart rate variability indices using random forest algorithm. The value of importance explains the degree of declining in the accuracy of data outside the bag when random noise is added. In theory, the sum of the importance value of all indicators is 1

TABLE 4 Results of 10-fold crossvalidation-based support vector machine classifier for classifying New York Heart Association classes I, II, and III groups

	Classification results									
	Class I			Class II			Class III			Average *Acc*
Fold	*Se* (%)	*Sp* (%)	*Acc* (%)	*Se* (%)	*Sp* (%)	*Acc* (%)	*Se* (%)	*Sp* (%)	*Acc* (%)	*Acc* (%)
1	68.4	97.3	92.9	48.7	92.8	79.6	91.4	57.0	77.6	83.4
2	76.1	96.6	94.2	52.1	93.9	81.5	91.2	60.3	79.2	85.0
3	64.7	97.6	92.5	54.2	92.6	80.8	90.2	58.6	77.1	83.5
4	79.4	96.9	94.2	57.6	93.1	82.7	90.3	66.4	80.7	85.9
5	71.4	97.5	93.4	52.2	95.2	82.0	93.5	60.0	79.4	85.0
6	58.8	96.9	89.5	48.8	92.3	79.2	90.3	54.4	74.8	81.2
7	71.7	98.4	94.1	56.6	93.7	82.5	92.6	63.1	80.2	85.6
8	64.7	97.0	92.0	56.4	94.2	82.3	92.0	60.8	78.9	84.4
9	70.9	97.9	93.2	48.6	90.8	77.9	89.0	57.4	75.5	82.2
10	69.5	97.9	93.1	50.5	93.3	80.4	92.0	58.6	77.9	83.8
Mean	69.6	97.4	92.9	52.6	93.2	80.9	91.3	59.7	78.1	84.0
SD	5.6	0.5	1.3	3.3	1.1	1.5	1.3	3.2	1.8	1.4

TABLE 5 Results of 10-fold crossvalidation-based classification and regression tree classifier for classifying New York Heart Association classes I, II, and III groups

	Classification results Class I			Class II			Class III			Average *Acc*
Fold	*Se* (%)	*Sp* (%)	*Acc* (%)	*Se* (%)	*Sp* (%)	*Acc* (%)	*Se* (%)	*Sp* (%)	*Acc* (%)	*Acc* (%)
1	60.6	97.3	91.9	53.8	88.5	78.1	86.8	58.2	75.5	81.8
2	59.7	95.7	90.2	51.2	91.8	79.0	89.3	56.5	75.8	81.7
3	51.8	97.7	90.3	55.0	88.8	72.3	87.2	55.9	74.3	79.0
4	52.6	97.9	91.1	57.9	89.5	80.9	88.1	58.2	77.0	83.0
5	53.7	97.8	89.7	58.0	91.9	80.8	89.8	58.8	76.3	82.3
6	59.0	96.1	90.3	55.9	91.7	80.4	88.3	58.7	76.0	82.2
7	58.4	96.0	89.4	53.3	90.8	78.3	87.8	57.1	74.1	80.6
8	54.4	96.3	89.1	55.2	87.8	77.0	85.6	57.5	73.3	79.8
9	59.1	96.5	89.5	53.7	87.6	77.9	85.8	59.1	74.9	80.8
10	55.8	96.0	90.2	60.1	89.2	80.7	87.0	62.3	77.6	82.8
Mean	56.5	96.7	90.2	55.4	89.8	78.5	87.6	58.2	75.5	81.4
SD	3.1	0.8	0.8	2.5	1.9	2.5	1.3	1.7	1.3	1.3

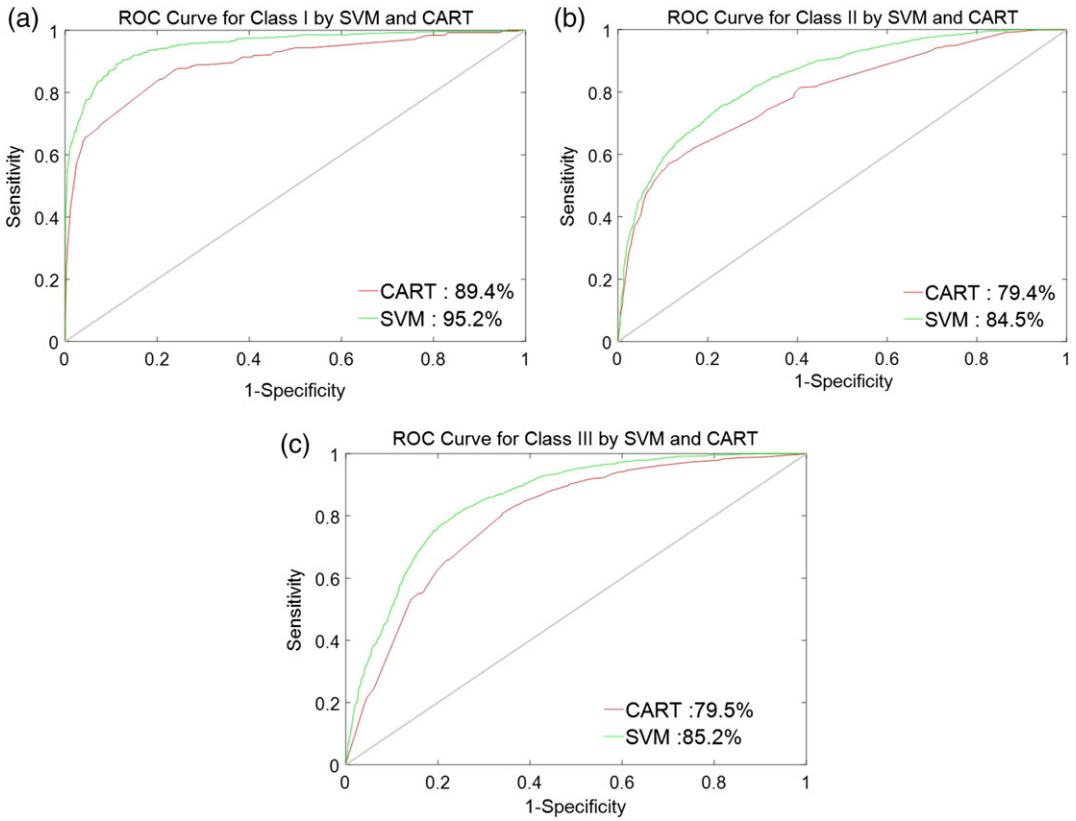

FIGURE 5 Results of receiver operating characteristic (ROC) curve analysis for evaluating the performance of the support vector machine (SVM) classifier and the classification and regression tree (CART) classifier for discriminating congestive heart failure patients with New York Heart Association functional classes I (a), II (b), and III (c)

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant 81871444 and the Primary Research & Development Plan of Jiangsu Province under Grant BE2017735. The authors thank the support from the Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY

The data used to support the findings of this study are available from the open-access MIT-BIH RR Interval Database.

ORCID


Chengyu Liu https://orcid.org/0000-0003-1965-3020

REFERENCES

Acharya, U. R., Joseph, K. P., Kannathal, N., Min, L. C., & Suri, J. S. (2007). Heart rate variability. In *Advances in cardiac signal processing* (pp. 121–165). New York: Springer.

Alpaydin, E. (2009). *Introduction to machine learning*. Boston: MIT Press.

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, *52*(4), 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015

Aronson, D., Mittleman, M. A., & Burger, A. J. (2004). Measures of heart period variability as predictors of mortality in hospitalized patients with decompensated congestive heart failure. *The American Journal of Cardiology*, *93*(1), 59–63. https://doi.org/10.1016/j.amjcard.2003.09.013

Bennett, J. A., Riegel, B., Bittner, V., & Nichols, J. (2002). Validity and reliability of the NYHA classes for measuring research outcomes in patients with cardiac disease. *Heart & Lung: The Journal of Acute and Critical Care*, *31*(4), 262–270. https://doi.org/10.1067/mhl.2002.124554

Berntson, G. G., & Stowell, J. R. (1998). ECG artifacts and heart period variability: Don't miss a beat! *Psychophysiology*, *35*(1), 127–132. https://doi.org/10.1111/1469-8986.3510127

Blanz, V., Schölkopf, B., Bülthoff, H., Burges, C., Vapnik, V., & Vetter, T. (1996). Comparison of view-based object recognition algorithms using realistic 3D models. Paper presented at the International Conference on Artificial Neural Networks.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brennan, M., Palaniswami, M., & Kamen, P. (2001). Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Transactions on Biomedical Engineering*, *48*(11), 1342–1347. https://doi.org/10.1109/10.959330

Camm, A., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., … Lombardi, F. (1996). 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. *Circulation*, *93*(5), 1043–1065.

Casolo, G., Balli, E., Taddei, T., Amuhasi, J., & Gori, C. (1989). Decreased spontaneous heart rate variability in congestive heart failure. *American Journal of Cardiology*, *64*(18), 1162–1167. https://doi.org/10.1016/0002-9149(89)90871-0

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Christou, N., & Dinov, I. D. (2011). Confidence interval based parameter estimation—A new SOCR applet and activity. *PLoS ONE*, *6*(5), e19178. https://doi.org/10.1371/journal.pone.0019178

Clearwater, S., & Stern, E. (1991). A rule-learning program in high energy physics event classification. *Computer Physics Communications*, *67*(2), 159–182. https://doi.org/10.1016/0010-4655(91)90014-C

Cohn, J. N. (1996). The management of chronic heart failure. *New England Journal of Medicine*, *335*(7), 490–498. https://doi.org/10.1056/NEJM199608153350707

Costa, M., Goldberger, A. L., & Peng, C.-K. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, *89*(6), 068102. https://doi.org/10.1103/PhysRevLett.89.068102

D'addio, G., Acanfora, D., Pinna, G., Maestri, R., Furgi, G., Picone, C., & Rengo, F. (1998). Reproducibility of short-and long-term Poincare plot parameters compared with frequency-domain HRV indexes in congestive heart failure. Paper presented at the Computers in Cardiology 1998.

Fonarow, G. C., Adams, K. F., Abraham, W. T., Yancy, C. W., Boscardin, W. J., & Committee, A. S. A. (2005). Risk stratification for in-hospital mortality in acutely decompensated heart failure: Classification and regression tree analysis. *Jama*, *293*(5), 572–580.

Franc, V., & Hlavác, V. (2002). Multi-class support vector machine. Paper presented at the Pattern Recognition, 2002. Proceedings. 16th International Conference on.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., … Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, *101*(23), e215–e220.

Graff, B., Graff, G., & Kaczkowska, A. (2012). Entropy measures of heart rate variability for short ECG datasets in patients with congestive heart failure. *Acta Physica Polonica B Proceedings Supplement*, *5*(1), 153–158. https://doi.org/10.5506/APhysPolBSupp.5.153

Hadase, M., Azuma, A., Zen, K., Asada, S., Kawasaki, T., Kamitani, T., … Matsubara, H. (2004). Very low frequency power of heart rate variability is a powerful predictor of clinical prognosis in patients with congestive heart failure. *Circulation Journal*, *68*(4), 343–347. https://doi.org/10.1253/circj.68.343

Ho, K. K., Moody, G. B., Peng, C.-K., Mietus, J. E., Larson, M. G., Levy, D., & Goldberger, A. L. (1997). Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation*, *96*(3), 842–848. https://doi.org/10.1161/01.CIR.96.3.842

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification.

İşler, Y., & Kuntalp, M. (2007). Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine*, *37*(10), 1502–1510. https://doi.org/10.1016/j.compbiomed.2007.01.012

Kamen, P., & Tonkin, A. M. (1995). Application of the Poincaré plot to heart rate variability: A new measure of functional status in heart failure. *Australian and New Zealand Journal of Medicine*, *25*(1), 18–26. https://doi.org/10.1111/j.1445-5994.1995.tb00573.x

Keerthi, S. S., & Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, *15*(7), 1667–1689. https://doi.org/10.1162/089976603321891855

Khaled, A. S., Owis, M. I., & Mohamed, A. S. (2006). Employing time-domain methods and poincaré plot of heart rate variability signals to detect congestive heart failure. *BIME Journal*, *6*(1), 35–41.

Lee, H. G., Noh, K. Y., & Ryu, K. H. (2007). Mining biosignal data: Coronary artery disease diagnosis using linear and nonlinear features of HRV. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lin, H.-T., & Lin, C.-J. (2003). A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Neural Computation*, *3*, 1–32.

Liu, C., & Gao, R. (2017). Multiscale entropy analysis of the differential RR interval time series signal and its application in detecting congestive heart failure. *Entropy*, *19*(6), 251.

Liu, G., Wang, L., Wang, Q., Zhou, G., Wang, Y., & Jiang, Q. (2014). A new approach to detect congestive heart failure using short-term heart rate variability measures. *PLoS ONE*, *9*(4), e93399. https://doi.org/10.1371/journal.pone.0093399

Malik, M. (1998). Heart rate variability. *Current Opinion in Cardiology*, *13*(1), 36–44. https://doi.org/10.1097/00001573-199801000-00006

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Masetic, Z., & Subasi, A. (2016). Congestive heart failure detection using random forest classifier. *Computer Methods and Programs in Biomedicine*, *130*, 54–64. https://doi.org/10.1016/j.cmpb.2016.03.020

Mehrabi, S., Maghsoudloo, M., Arabalibeik, H., Noormand, R., & Nozari, Y. (2009). Application of multilayer perceptron and radial basis function neural networks in differentiating between chronic obstructive pulmonary and congestive heart failure diseases. *Expert Systems with Applications*, *36*(3), 6956–6959. https://doi.org/10.1016/j.eswa.2008.08.039

Melillo, P., De Luca, N., Bracale, M., & Pecchia, L. (2013). Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. *IEEE Journal of Biomedical and Health Informatics*, *17*(3), 727–733. https://doi.org/10.1109/JBHI.2013.2244902

Nolan, J., Batin, P. D., Andrews, R., Lindsay, S. J., Brooksby, P., Mullen, M., … Fox, K. A. (1998). 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). *Circulation*, *98*(15), 1510–1516. https://doi.org/10.1161/01.CIR.98.15.1510

Pecchia, L., Melillo, P., & Bracale, M. (2011). Remote health monitoring of heart failure with data mining via CART method on HRV features. *IEEE Transactions on Biomedical Engineering*, *58*(3), 800–804. https://doi.org/10.1109/TBME.2010.2092776

Pecchia, L., Melillo, P., Sansone, M., & Bracale, M. (2011). Discrimination power of short-term heart rate variability measures for CHF assessment. *IEEE Transactions on Information Technology in Biomedicine*, *15*(1), 40–46. https://doi.org/10.1109/TITB.2010.2091647

Ponikowski, P., Anker, S. D., Chua, T. P., Szelemej, R., Piepoli, M., Adamopoulos, S., … Wrabec, K. (1997). Depressed heart rate variability as an independent predictor of death in chronic congestive heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. *The American Journal of Cardiology*, *79*(12), 1645–1650. https://doi.org/10.1016/S0002-9149(97)00215-4

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), H2039–H2049.

Rutkowski, L., Jaworski, M., Pietruczuk, L., & Duda, P. (2014). The CART decision tree for mining data streams. *Information Sciences*, *266*, 1–15. https://doi.org/10.1016/j.ins.2013.12.060

Senni, M., Tribouilloy, C. M., Rodeheffer, R. J., Jacobsen, S. J., Evans, J. M., Bailey, K. R., & Redfield, M. M. (1998). Congestive heart failure in the community: A study of all incident cases in Olmsted County, Minnesota, in 1991. *Circulation*, *98*(21), 2282–2289. https://doi.org/10.1161/01.CIR.98.21.2282

Signorini, M., Ferrario, M., Marchetti, M., & Marseglia, A. (2006). Nonlinear analysis of heart rate variability signal for the characterization of cardiac heart failure patients. Paper presented at the Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE.

Stein, P. K., Bosner, M. S., Kleiger, R. E., & Conger, B. M. (1994). Heart rate variability: A measure of cardiac autonomic tone. *American Heart Journal*, *127*(5), 1376–1381. https://doi.org/10.1016/0002-8703(94)90059-0

Stein, P. K., Rich, M. W., Rottman, J. N., & Kleiger, R. E. (1995). Stability of index of heart rate variability in patients with congestive heart failure. *American Heart Journal*, *129*(5), 975–981. https://doi.org/10.1016/0002-8703(95)90119-1

Sztajzel, J. (2004). Heart rate variability: A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly*, *134*(35–36), 514–522. 2004/35/smw-10321

Zhao, L., Wei, S., Zhang, C., Zhang, Y., Jiang, X., Liu, F., & Liu, C. (2015). Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects. *Entropy*, *17*(9), 6270–6288. https://doi.org/10.3390/e17096270

AUTHOR BIOGRAPHIES

Zhaohui Qu is a senior student majoring in Instrument Science and Engineering in Southeast University, China. Now, he is working as a research assistant in Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab, Southeast University, China. His research interests include physiological signal processing and machine learning.

Qianwen Liu is a senior student majoring in Instrument Science and Engineering in Southeast University, China. Now, she is working as a research assistant in Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab, Southeast University, China. Her research interests include biomedical signal processing and big data processing.

Chengyu Liu received his B.S. and Ph.D. degrees in Biomedical Engineering from Shandong University, China, in 2005 and 2010, respectively. Dr. Liu has completed the Postdoctoral trainings at Shandong University in China (2010–2013), Newcastle University in the United Kingdom (2013–2014), and Emory University in the United States (2015–2017). He is currently a Professor at School of Instrument Science and Engineering, Southeast University, China. He is also the Director of Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab. Dr. Liu was the PI on 10+ awarded grants. He has published more than 150 papers. He is now a Federation Journal Committee Member of International Federation for Medical and Biological Engineering (IFMBE) and the chair for China Physiological Signal Challenge. His research topics include wearable mHealth and intelligent monitoring, machine learning and big data processing for cardiovascular signals, device development, and sleep and emotion monitoring.

How to cite this article: Qu Z, Liu Q, Liu C. Classification of congestive heart failure with different New York Heart Association functional classes based on heart rate variability indices and machine learning. *Expert Systems*. 2019;e12396. https://doi.org/10.1111/exsy.12396