Feature Selection Scheme Based on Multi-time-scales for Analyzing Congestive Heart Failure

Chunyuan Wang1,2, Yatao Zhang3, Xinge Jiang4, Feifei Liu5, Zhimin Zhang6 and Shoushui Wei7

1 School of Control Science and Engineering, Shandong University, Jinan, China
2 Engineering Research Center, Suzhou Institute of Biomedical Engineering and Technology Chinese Academy of Sciences, Suzhou, China
3 School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China
4 School of Information Science and Electrical Engineering, Shandong Jiaotong University, Jinan, China
5 School of Science, Shandong Jianzhu University, Jinan, China
6 Science and Technology on Information Systems Engineering Laboratory, The 28th Research Institute of CETC, Nanjing, China
7 School of Instrument Science and Engineering, China Southeast University, Nanjing, China

Abstract

This paper proposed a feature selection method combined with multi-time-scales analysis and heart rate variability (HRV) analysis for middle and early diagnosis of congestive heart failure (CHF). In previous studies regarding the diagnosis of CHF, researchers have tended to increase the variety of HRV features by searching for new ones or to use different machine learning algorithms to optimize the classification of CHF and normal sinus rhythms subject (NSR). In fact, the full utilization of traditional HRV features can also improve classification accuracy. The proposed method constructs a multi-time-scales feature matrix according to traditional HRV features that exhibit good stability in multiple time-scales and differences in different time-scales. The multi-scales features yield better performance than the traditional single-time-scales features when the features are fed into a support vector machine (SVM) classifier, and the results of the SVM classifier exhibit a sensitivity, a specificity, and an accuracy of 99.52%, 100.00%, and 99.83%, respectively. These results indicate that the proposed feature selection method can effectively reduce redundant features and computational load when used for automatic diagnosis of CHF.

Keywords: feature selection, multi-time-scales analysis, congestive heart failure, heart rate variability

1. Introduction

Cardiovascular diseases are the leading cause of death worldwide, and heart failure, also known as congestive heart failure (CHF), is a highly lethal type of cardiovascular disease. Heart rate variability (HRV) analysis is an effective tool to evaluate the overall heart health and the state of the autonomic nervous system, which reflects the regulation of cardiac activity. In recent years, several studies have been conducted on the use of electrocardiogram (ECG) and HRV analysis for the diagnosis of CHF. Asyali [1] determined nine optimal features within the time and frequency domains for CHF diagnosis. These nine HRV features were then fed into the linear discriminant analysis and Bayesian classifier for CHF diagnosis. The SDNN features in the time domain were concluded to have a high-resolution ability for CHF and normal sinus rhythms subject (NSR) classification. Ali Narin et al. [2] used several features to determine the best combination of features to distinguish CHF and NSR. Furthermore, wavelet packet transform-based frequency-
domain features and several nonlinear parameters were used in addition to standard HRV features. The backward elimination method was used to select 27 HRV feature combinations to obtain an ideal classification effect using a support vector machine (SVM) classifier. Cornforth et al. [3] used the best Renyi entropy combined with traditional time-domain features for CHF identification; the accuracy obtained was higher than that obtained when solely time-domain features were used. Kumar et al. [4] used the flexible analytic transform method to decompose HRV signals into sub-band signals and then extracted 20 features using accumulated fuzzy entropy and accumulated permutation from different frequency scales. The Bhattacharyya ranking method and least squares SVM (LS-SVM) were used to construct the CHF autonomous diagnosis system for obtaining the optimal results. Li et al. [5] proposed a convolutional neural network and distance distribution matrix in the entropy calculation method to classify CHF and NSR, and they proved the effectiveness of the combination. Yalcin Isler et al. [6] selected features in the time, frequency, and nonlinear domains according to the significance level for feature combination. K-nearest neighbors, linear discriminant analyses, multilayer perceptron, SVMs, and a radial basis function artificial neuronal network were used for testing. Finally, a high-accuracy automatic diagnosis CHF system was proposed.

In previous studies on the diagnosis of CHF, researchers have tended to increase the variety of features by searching for new ones or to use different machine learning algorithms to optimize the classifications. These traditional methods will lead to a sharp increase in the number of features. Furthermore, as is well known, the quality of the selected features will affect the classification results. Therefore, if the number of selected features is extremely high, among which low-quality and redundant features exist, not only does the operation time and cost increase, but the classification accuracy also decreases. In contrast, the full utilization of traditional HRV features can improve the accuracy of the classifier. In general, the value range of a feature should be an interval. However, when calculating the feature of a subject, the feature is often represented as a number, which is actually inaccurate. For ensuring the validity of the feature, an appropriate method to transform the feature into the feature matrix can effectively improve the accuracy of the classifier. In this study, a single feature is extended to a feature matrix using a multi-time-scales analysis method.

Multi-time-scales analysis is widely used in the signal analysis of natural time series. Peng et al. [7] demonstrated that multi-time-scales analysis could be used for neurophysiological control mechanisms using heart rate and gait regulation as features. The time-scale of the features can be used to develop early warning systems for subjects with various pathological characteristics, particularly for subjects at high risk of sudden death, that is, CHF. This study laid the foundation for the classification of CHF using multi-time-scales analysis. Chladekova et al. [8] proposed the time irreversibility of new methods using HRV and blood pressure variability to evaluate the cardiovascular regulating mechanism. They used three different time irreversibility indices—Porta’s, Guzik’s, and Ehler’s indices derived from data segments containing 1000 beat-to-beat intervals on four time-scales. Their study helped identify a clear link between the temporal irreversibility of HRV features and the autonomic nervous system. Costa et al. [9] found that multi-scales entropy and multi-scale time irreversibility could extract information from the time series between beats. Their study suggested that multi-time-scales analysis has a reference value for the diagnosis of heart disease. Hu et al. [10] continued and completed the work of Costa et al. by selecting the optimal time-scale for the features in the analysis of HRV for the diagnosis of CHF. In addition, new nonstandard features were constructed based on feature variation trends in multi-time-scales analysis to optimize the classification effect of patients with CHF and NSR.

This paper proposed a feature selection method using multi-time-scales analysis. For screening suitable time-scales and high-quality features, a feature matrix combined with different time-scales features and HRV features was constructed and input into an SVM classifier. Compared with the traditional analysis method, the number of features is pruned to avoid possible faults caused by low-quality features and redundant features in the traditional method. In fact, a better classification performance can be guaranteed while the number of features is optimized.


2. Materials and Method

2.1 Dataset

The complex physiological signal database PhysioBank [11] was used in this study. The Normal Sinus Rhythm RR Interval Database (nsr2db) was used as the sample source of a normal heart rate, which included 54 patients with long-term ECG signals of a normal rhythm collected from 30 males (ranging in age from 28.5 to 76 years) and 24 females (ranging in age from 58 to 73 years old). The age distribution of the subjects was mostly concentrated in the middle and old age, and the age distribution was not used as a classification feature in this study. The CHF RR Interval Database (chf2db), as a sample source of patients with CHF, contains 29 long-term ECG signals accompanying CHF (the sex ratio is unknown, and the age distribution is between 34 and 79 years), including 4 NYHA I patients, 8 NYHA II patients, and 17 NYHA III patients. NYHA is a rating for the severity of CHF [12]. In the two databases, the duration of all RR interval sequences was approximately 24 h, and the sampling rate was 128 Hz. The beat annotations were obtained by automated analysis with manual review and correction. The sample ratio of the two types were ranged between 1:1 and 1:2, and therefore, any sample imbalance was considered as negligible.

In most cases, RR intervals of less than 0.4 s may indicate that an R peak was incorrectly detected within a normal RR interval, while > 2.0 s may indicate that an R peak was missed between two normal RR intervals [10]. Therefore, RR intervals less than 0.4 s or greater than 2.0 s were deleted from the original data. In this study, each signal should be segmented to calculate the HRV features at different time-scales. After deleting the above invalid data, we obtained signals where the length of each segment is approximately 18 h to 24 h, among which the total time length of the deleted invalid data accounts for approximately 3% of the total time length of all data.

2.2 Method

The proposed feature selection method consists of three steps. First, the traditional HRV features are found and divided into multiple time-scales. Second, the significant differences between the feature data of the CHF and NSR groups were analyzed. Features with significant differences and multi-time-scales stability were selected. Finally, a multi-time-scales feature matrix input classifier is constructed by determining the features that meet the requirements. The basic flow of the feature selection is shown in Fig. 1.

2.2.1 HRV features. In this study, nine traditional features including the time, frequency, and nonlinear domains were first calculated [10]. Time domain features include MEAN, SDNN, and RMSSD [13], where the MEAN is the average value of the RR interval, the SDNN was used to evaluate the overall variability of the heart rates, and the RMSSD was used to evaluate the short-term variability of the heart rates. The frequency domain features include LF_n, HF_n, and Ratio-LH [14], where LF_n is the low-frequency changes of the heart rates, HF_n is the high-frequency changes of the heart rates [15], and Ratio-LH is the ratio of the low-frequency and high-frequency. Nonlinear domain features

![Fig. 1 Basic flow chart of the feature selection](image-url)
2.2.2 Dividing time-scales. In this study, the difference between NSR and CHF was analyzed using the multi-time-scales analysis method to calculate nine features in the RR interval sequence with seven time-scales (i.e., 5 min, 10 min, 30 min, 1 h, 2 h, 5 h, and 10 h) according to the power of two and the habits of the dividing timing scale [10]. Because there exist samples with data lengths less than 20 h after preprocessing, the maximum time-scale of this study was selected to be 10 h. These nine features were extended to the seven time-scales by using the multi-time-scales analysis method, and the statistical differences of the HRV features between the two groups of CHF and NSR were obtained on seven time-scales. In the analysis, the average value of the calculated value of the same feature in all segments was taken as the final calculation result of the feature in the sample.

2.2.3 Feature selection. After obtaining the HRV measurement values of NSR and CHF on seven time-scales, the independent double-sample T-test was used to determine whether there were significant differences between the two groups. Before the t test statistics, the Kolmogorov–Smirnov test was used to determine if the HRV features in both groups were normally distributed [19]. MATLAB (Ver. 2014a, MathWorks) was used for all statistical analyses, with the significance level (α) set to 0.05.

Table 1 Difference levels of the HRV features between CHF and NSR on the seven time scales

| p-value | 5 min | 10 min | 30 min | 1 h | 2 h | 5 h | 10 h |
|---------|-------|--------|--------|-----|-----|-----|------|
| MEAN    | 6.75E-06 | 6.63E-06 | 6.09E-06 | 5.41E-06 | 3.92E-06 | 3.05E-06 | 2.92E-06 |
| SDNN    | 0.957 | 0.644 | 0.206 | 7.94E-02 | 1.73E-02 | 1.57E-03 | 6.01E-06 |
| RMSSD   | 1.67E-03 | 1.43E-03 | 1.25E-03 | 1.22E-03 | 1.29E-03 | 1.43E-03 | 1.72E-03 |
| LF_n    | 1.90E-10 | 1.14E-10 | 5.48E-11 | 4.71E-11 | 4.02E-11 | 2.95E-11 | 4.52E-11 |
| HF_n    | 1.90E-10 | 1.14E-10 | 5.48E-11 | 4.71E-11 | 4.02E-11 | 2.95E-11 | 4.52E-11 |
| Ratio-LH | 1.26E-11 | 1.59E-11 | 1.06E-11 | 9.04E-12 | 1.40E-11 | 1.25E-11 | 1.94E-11 |
| VAI     | 2.47E-02 | 2.46E-02 | 2.47E-02 | 2.45E-02 | 2.49E-02 | 2.57E-02 | 2.81E-02 |
| VLI     | 2.58E-03 | 1.06E-03 | 1.82E-04 | 8.56E-05 | 1.22E-05 | 3.51E-06 | 2.40E-08 |
| SmpE    | 1.43E-04 | 1.31E-03 | 0.120 | 0.881 | 0.164 | 2.60E-02 | 4.73E-04 |
Fig. 2 Statistical bar graph of the mean values and standard deviations of the nine HRV features in NSR and CHF at seven time-scales (the blue line is NSR, and red CHF).

Table 1 indicates that SDNN and SmpE have a poor significance in distinguishing the CHF and NSR on partial time-scales (the feature data with poor significance are in bold in Table 1). All the other features have good significance at different time scales.

Table 1 and Fig. 2 show that SDNN, Ratio-LH, VLI, and SmpE have changing trends as the time scale changes (the dashed line in the figure is the trend line). If these features are used to construct the feature matrix, their changing trend will cause obvious interference. In contrast, the MEAN, RMSSD, LF, and VAI are intuitively stable at all time-scales, and significant differences between the groups NSR and CHF exist.

2.2.4 Feature matrix. With the above five features MEAN, RMSSD, LF, HF, and VAI, the multi-time-scales feature matrix is defined as

\[ I = [a_{nt}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & a_{n5} & a_{n6} & a_{n7} \end{bmatrix} \]  

(1)

where \( I \) represents the feature matrix, \( n \) represents the number of samples, \( t \) represents the time scale, and \( a_{nt} \) represents the feature value under the time scale of \( t \) of the \( n \) samples.

2.2.5 Classification method. HRV features were used as the feature space; an SVM classifier based on the Gaussian kernel of RBF (RBF-SVM) [20] was used to establish an automatic diagnosis model of heart failure. Furthermore, a grid search algorithm was applied to determine the best combination of penalty coefficient and function parameters. The two parameters ranged from \( 10^{-5} \) to \( 10^5 \). There are \( 121 \) different combinations. The model was built on the Spyder platform using the Scikit-Learn machine learning library (Ver. 0.19.1) in Python (Ver. Python 3.6).
The SVM model performance was evaluated in terms of sensitivity (Se), specificity (Sp), and accuracy (Acc):

\[
Se = \frac{TP}{TP + FN}
\]  
\[
Sp = \frac{TN}{FP + TN}
\]  
\[
Acc = \frac{TP + TN}{TP + FN + FP + TN}
\]  

where TP, FP, FN, TN denote the number of true positives, false positives, false negatives and true negatives in confusion matrix. The used dataset was evaluated by the 10-fold cross validation method to assess the generalization ability of the model, with the average value of the ten results as the final evaluation result. In section 3, Se, Sp, and Acc are all the classification results of the NSR and CHF dichotomies. The preprocessing and machine learning methods for each model used for comparison are consistent and will not be described in the next section.

### 3. Results

Different single time-scale features were used as an input to the SVM classifiers for training, and the results of different time-scales were compared.

Table 2 Comparison of nine features between single-time scale prediction model and proposed prediction model

| Time scale | Se      | Sp      | Acc     |
|------------|---------|---------|---------|
| 5 min      | 70.00   | 98.33   | 88.89   |
| 10 min     | 80.00   | 96.67   | 91.11   |
| 30 min     | 83.33   | 98.33   | 93.33   |
| 1 h        | 83.33   | 98.67   | 93.89   |
| **2 h**    | **86.67** | **98.33** | **94.44** |
| 5 h        | 86.67   | 91.67   | 90.00   |
| 10 h       | 83.67   | 96.67   | 92.44   |

Table 2 summarizes the performance of the classification model based on SVM machine learning using all nine features. In the 2 h time scale, the classification was the best. The classification was performed with a sensitivity, a specificity, and an accuracy of 86.67%, 98.33%, and 94.44%, respectively. The results of this study are similar to those of [10]. Because the time scale analysis method is the same as the classification method, the authenticity of the results is also verified.

Table 3 Optimal classification results of the single time scale and classification results using the proposed multiple time scale feature matrix

|          | Se    | Sp    | Acc    |
|----------|-------|-------|--------|
| 2 h      | 86.67 | 98.33 | 94.44  |
| Proposed | 99.52 | 100.00| 99.83  |

Table 3 summarizes the performance of the SVM after constructing the feature matrix using the five features. The classification was performed with a sensitivity, a specificity, and an accuracy of 99.52%, 100.00% and 99.83%, respectively. Compared with the optimal classification results in Table 4, despite the types of features that were reduced, the classification results significantly improved after strengthening with the use of the high-quality features.

In this study, the leave-one-out-cross-validation (LOOCV) method was employed to verify the performance of the classification model. The LOOCV is also one of the cross-validation methods: leaving out one sample without expansion and training and using the remaining samples for building the model after sample expansion. Seven time scale features of the retained samples were fed into the model for testing, and the final classification results of the samples were voted on seven times. The specific method is shown in Fig. 3.
The experimental results showed that all 54 healthy subjects and 29 patients with heart failure were successfully distinguished.

Table 4 Literature comparison

| Author          | Year | Data (Physionet) | Number of features | Timescale       | Classifier   | Se  | Sp  | Acc  |
|-----------------|------|------------------|--------------------|-----------------|--------------|-----|-----|------|
| Narin et al.[2] | 2014 | nsr2db(54)       | 27                 | 5 min           | SVM          | 82.75 | 96.29 | 91.56 |
| David et al.[3] | 2016 | nsrdb(18)        | 4                  | 1000 R-R intervals | KNN         | 80.00 | 94.40 | 87.9  |
| Kumar et al.[4] | 2017 | nsrdb(18)        | 18                 | 500 samples     | LS-SVM       | 98.07 | 98.33 | 98.21 |
| Wang et al.[19]| 2018 | nsr2db(52)       | 9                  | 5 min           | SVM          | 91.31 | 90.04 | 90.95 |
| Isler et al.[6] | 2019 | nsr2db(54)       | 34                 | 5 min           | Three-stage classifier | 100.0 | 98.10 | 98.80 |
| Hu et al.[10]   | 2019 | nsr2db(54)       | 3                  | 5 min to 10 h   | SVM          | 93.33 | 98.33 | 96.67 |
| Proposed Method |      | nsr2db(54)       | 5                  | 5 min to 10 h   | SVM          | **99.52** | **100.0** | **99.83** |

Table 4 compares the results obtained using the literature model results and proposed model to distinguish CHF from NSR using HRV measures. HRV analysis is currently used in the diagnosis of heart failure in literature; those generally have more features or more complex classifiers. Through using the method of constructing a multi-time-scales feature matrix and filtering for high-quality features, fewer feature types are entered into the classifier. Good classification results have been obtained; therefore, this method can effectively save time and power consumption.

4. Discussion

4.1 Time scale selection

Theoretically, more data feature matrices can be obtained by selecting more time-scales, and the classification results should be better. In this study, when the time scale was divided to be extremely fine, the eigenvalues of adjacent time-scales were very similar. This does not help improve the classification results; further, the significance of constructing the feature matrix is lost. Different time-scales are selected for comparison to determine the appropriate multi-time scale model matrix.

Based on previous studies [10], the best classification results were achieved at the 2 h time scale. Owing to the limited length of the data, the problem of insufficient data volumes will be encountered when selecting a long time scale. Therefore, when the time scale is increased, an expansion of 2 h in the medium time scale is preferred. Namely, 5 min, 10 min, 30 min, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, 3.5 h, 4.5 h, 5 h, and 10 h, among which the 3, 5, 7, 9, 11, and 13 time-scales were selected for comparison. The specific rules are listed in Table 5, where the populated part of the table represents the selected time scales.
Table 5 Time scale used by each model

| Expansion multiple | 3   | 5   | 7   | 9   | 11  | 13  |
|--------------------|-----|-----|-----|-----|-----|-----|
| 5 min              |     |     |     |     |     |     |
| 10 min             |     |     |     |     |     |     |
| 30 min             |     |     |     |     |     |     |
| 1 h                |     |     |     |     |     |     |
| 1.5 h              |     |     |     |     |     |     |
| 2 h                |     |     |     |     |     |     |
| 2.5 h              |     |     |     |     |     |     |
| 3 h                |     |     |     |     |     |     |
| 3.5 h              |     |     |     |     |     |     |
| 4 h                |     |     |     |     |     |     |
| 4.5 h              |     |     |     |     |     |     |
| 5 h                |     |     |     |     |     |     |
| 10 h               |     |     |     |     |     |     |

Fig. 4 shows that the classification result is close to the optimal result when the scale value of the time-scales is seven. Therefore, time-scale seven was selected to construct the feature matrix in this study.

Fig. 4 Comparison of expanding multiple results of different time-scales (the red line is Acc; blue line, Se; and green line, Sp)

4.2 Rationality analysis of feature selection

The rationality of the five selected features is further analyzed. First, the multi-time-scales stability of the features was verified by calculating the Pearson correlation coefficient between the features and time scale. Second, the differences in the features at different time-scales are analyzed.

4.2.1 Analysis of feature stability. In this study, the Pearson correlation coefficient was used to verify the correlation between the selected features and time scale. The correlation coefficient is defined as follows:

$$r = \frac{N \sum at - \sum a \sum t}{\sqrt{N \sum a^2 - (\sum a)^2} \sqrt{N \sum t^2 - (\sum t)^2}}$$  \hspace{1cm} (5)

where $t$ represents the aforementioned time scales, $a$ is the corresponding feature in the time scale of $t$, and $r$ is the correlation coefficient. The closer the correlation coefficient is to 1 or -1, the stronger is the correlation. The closer the correlation coefficient is to 0, the weaker is the correlation. When the absolute value of the correlation coefficient is less than 0.2, the two variables are considered to have no or very weak correlation.

| CHF      | RMSSD | LF$_a$ | HF$_a$ | VAI |
|----------|-------|--------|--------|-----|
| 0.028    | 0.070 | 0.040  | 0.040  | 0.003 |
| NSR      | 0.027 | 0.061  | 0.024  | 0.024 | 0.010 |

Table 6 lists that the Pearson correlation coefficients of the five features selected from the CHF and NSR are far less than 0.2. It indicates that the five features have no obvious change trend with the change of time-scale. Thus, the values of the five features are relatively stable at different time scales.

4.2.2 Analysis of feature difference. The method of constructing the feature matrix is similar to resampling. The features of different time-scales are similar to resampling sample sets. One problem that the resampling method avoids is the duplication of samples. Simply copying a sample is not valid for a classifier. Similar to the resampling method, the proposed method is necessary to ensure the difference in different time scales. The maximum difference of the same sample at different time scales is compared with the
minimum difference between different samples sorted by the size of the features. We compare to verify whether there is an order of magnitude difference between the two (The order of magnitude is base 10). If the differences between the two groups remain on the same order of magnitude, then the selected features are considered to be different on different time scales.

\[
S_{\text{mtr}} = E \left( \sum \left( \max(a_{i,t} | 5 \leq t \leq 10h) - \min(a_{i,t} | 5 \leq t \leq 10h) \right) \right)
\]

\[
S_{\text{iter}} = E \left( \sum \left( \max(a_{i,t}) - \min(a_{i,t}) \right) \right)
\]

where \( n \) represents the number of samples, \( t \) represents the time scale, \( a_{i,t} \) represents the feature value under the time scale of \( t \) of the \( n \) samples, \( S_{\text{mtr}} \) represents the average value of the maximum difference in all time scales of a feature in the \( n \) samples, and \( S_{\text{iter}} \) represents the average value of the minimum difference among all time scales of a feature in all samples.

Table 7 Differences of the intra-group and inter-group features were compared between patients with CHF and NSR.

| Feature Type | CHF | NSR |
|--------------|-----|-----|
|               | \( S_{\text{mtr}} \) | \( S_{\text{iter}} \) | \( S_{\text{mtr}} \) | \( S_{\text{iter}} \) |
| MEAN          | 6.48 | 11.40 | 17.28 | 8.06 |
| RMSSD         | 7.44 | 5.58  | 4.95  | 2.62 |
| LF_n          | 0.04 | 0.02  | 0.04  | 0.01 |
| HF_n          | 0.04 | 0.02  | 0.04  | 0.01 |
| VAI           | 0.06 | 0.13  | 0.03  | 0.05 |

Table 7 summarizes the differences between \( S_{\text{mtr}} \) and \( S_{\text{iter}} \) in the five features of the CHF and NSR and remained on the same order of magnitude, which indicates that the values of these five features are different at different time scales.

The above two steps help determine that the five features of MEAN, RMSSD, LF_n, HF_n, and VAI meet the conditions of the proposed method.

4.3 Comparing the two features schemes

The proposed method was first compared with the first feature scheme, namely five features under seven time-scales (35 features), which were directly input into the SVM classifier. The aim of this comparison is to evaluate whether the proposed method cannot be utilized to improve the number of features for classification performance.

Another classic resampling method, the synthetic minority oversampling technique (SMOTE), is based on the concept of creating a new minority sample [22]. Given that SMOTE can only use a single time scale, SMOTE chose the 2 h time scale with the best classification effect, using nine features to construct the model. The SMOTE method was used to expand the quantity from 1 to 7 times to compare with the method proposed in this study and choose \( k \) close to 3. The accuracy of the classifier was the best when the expansion multiple was 6.

Table 8 Comparison between the two prediction models and the prediction model in this study.

| Method            | Se   | Sp   | Acc  |
|-------------------|------|------|------|
| 35 Features       | 83.33| 96.67| 92.22|
| SMOTE             | 97.06| 98.48| 98.00|
| Proposed Method   | 99.52| 100.00| 99.83|

Table 7 indicates that compared with the prediction model using 35 features, the proposed method has been greatly improved in the three performance indicators. Simultaneously, no significant benefit is observed when the feature types are increased simply via multi-time scale analysis. Conversely, SMOTE can also achieve good classification results. However, using SMOTE, the sample expands by several times, causing the inevitable over-fitting problem. In general, using SMOTE can expand the sample by 2 to 3 times at most. The sample size was expanded by 6 times just to test the proposed method; this expansion is not practically significant. However, the features in the feature matrix with multiple time scales are all real eigenvalues and no overfitting problem caused by the large number of synthetic virtual features was observed. From this perspective, the method of this study is better than resampling.

5. Conclusion

This paper proposed a method of heart failure diagnosis using an SVM classifier by constructing a multi-time-scales feature matrix. This is different from the current method of increasing HRV feature types and using multiple classifiers.
The feature matrix was constructed using a multi-time-scales analysis, and the classification was optimized by screening the stable and high-quality features on different time scales. When selecting seven time-scales, the classifier has the best recognition effect with a sensitivity, a specificity, and an accuracy of 99.52%, 100.00%, and 99.83%, respectively. This method of using a multi-time scale feature matrix can potentially be applied for other biomedical signal processing scenarios, particularly those with an insufficient sample size.

In this study, the features that showed an evident change trend with time-scales were removed. However, a few laws hidden exist in the variation trend of these features, which need to be further studied.

Acknowledgements

We thank the MIT-BIH database for providing the invaluable data used in this study.

Declaration of Competing Interest

All authors have no competing interests to declare.

References

[1] Asyali M H 2003 Discrimination power of long-term heart rate variability measures, in International Conference of the IEEE Engineering in Medicine & Biology Society.
[2] Narin A, Isler Y and Ozer M 2014 Computers in Biology and Medicine 45 72–9 [DOI: 10.1016/j.compbiomed.2013.11.016] [PubMed: 24480166]
[3] Cornforth D and Jelinek H F 2016 Detection of Congestive Heart Failure Using Renyi Entropy, in Computing in Cardiology Conference.
[4] Kumar M, Pachori R B and Acharya U R 2017 Entropy 19 92 [DOI: 10.3390/e19030092]
[5] Yaowei L, Yao Z, Lina Z, Yang Z, Chengyu L, Li Z, Liuxin Z, Zhensheng L, Binhua W and Eky N 2018 IEEE Access 1–1.
[6] Isler Y, Narin A, Ozer M and Perc M 2019 Chaos Solitons & Fractals 118 145–51 [DOI: 10.1016/j.chaos.2018.11.020]
[7] Peng C K, Hausdorff J M, Havlin S, Mietus J E, Stanley H E and Goldberger A L 1998 Physica A: Statistical Mechanics and its Applications 249 491–500 [DOI: 10.1016/s0378-4371(97)00508-6] [PubMed: 11541513]
[8] Chladekova L, Czippelova B, Turianikova Z, Tonhajzerova I, Calkovska A, Baumert M and Javorka M 2012 Physiological Measurement 33 1747–56 [DOI: 10.1088/0967-3334/33/10/1747] [PubMed: 23010992]
[9] Costa M D, Peng C K and Goldberger A L 2008 Cardiovascular Engineering 8 88–93 [DOI: 10.1007/s10558-007-9049-1] [PubMed: 18172763]
[10] Hu B, Wei S, Wei D, Zhao L, Zhu G and Liu C 2019 IEEE Access 7 17862–71 [DOI: 10.1109/ACCESS.2019.2895998]
[11] Goldberger AL, et al. 2000 Circulation 101 215–20.
[12] Conditions, N.C.C.F.C., Chronic Heart Failure: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care 2003 (Royal College of Physicians).
[13] Brennan M, Palaniswami M and Kamen P 2001 IEEE Transactions on Biomedical Engineering 48 1342–7 [DOI: 10.1109/10.959330] [PubMed: 11686633]
[14] Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim C M and Suri J S 2006 Medical & Biological Engineering & Computing 44 1031–51 [DOI: 10.1007/s11517-006-0119-0] [PubMed: 17111118]
[15] Malik M and Terrace C 1996 European Heart Journal 17 354–81 [DOI: 10.1093/oxfordjournals.eurheartj.a014868] [PubMed: 8737210]
[16] Ruan X, Liu C, Liu C, Wang X and Li P Automatic Detection of Atrial Fibrillation Using R-R Interval Signal, in Biomedical Engineering and Informatics 2011.
[17] Richman J S and Moorman J R 2000 American Journal of Physiology-Heart & Circulatory Physiology 278 H2039–49 [DOI: 10.1152/ajpheart.2000.278.6.H2039] [PubMed: 10843903]
[18] Alangari H M and Sahakian A V 2007 IEEE Transactions on Biomedical Engineering 54 1900–4 [DOI: 10.1109/TBME.2006.889772] [PubMed: 17926691]
[19] Wang Y, Wei S, Zhang S, Zhang Y, Zhao L, Liu C and Murray A 2018 Biomedical Signal Processing and Control 42 30–6 [DOI: 10.1016/j.bspc.2018.01.001]
[20] Duda R O, Hart P E and Stork D G Linear Discriminant Functions.
[21] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R and Dubourg V 2011 Journal of Machine Learning Research 12 2825–30.
[22] Chawla N V, Bowyer K W, Hall L O and Kegelmeyer W P 2002 Journal of Artificial Intelligence Research 16 321–57 [DOI: 10.1613/jair.953].