

Article

Multiscale Entropy Analysis with Low-Dimensional Exhaustive Search for Detecting Heart Failure

Hsuan-Hao Chao ¹, Chih-Wei Yeh ¹, Chang Francis Hsu ¹, Long Hsu ¹ and Sien Chi ^{2,*} ¹ Department of Electrophysics, National Chiao Tung University, Hsinchu 30010, Taiwan² Department of Photonics, National Chiao Tung University, Hsinchu 30010, Taiwan

* Correspondence: schi@mail.nctu.edu.tw; Tel.: +886-3-5731824

Received: 3 July 2019; Accepted: 21 August 2019; Published: 24 August 2019



Abstract: Multiscale entropy (MSE) is widely used to analyze heartbeat signals. Even though cardiologists do not use MSE to diagnose heart failure at present, these studies are of importance and have potential clinical applications. In previous studies, MSE discrimination between old congestive heart failure (CHF) and healthy individuals has remained controversial. Few studies have been published on the discrimination between them, using only MSE with machine learning for automatic multidimensional analysis, with reported testing accuracies of less than 86%. In this study, we determined the optimal MSE scales for discrimination by using a low-dimensional exhaustive search along with three classifiers—linear discriminant analysis (LDA), support vector machine (SVM), and k-nearest neighbor (KNN). In younger people (<55 years), the results showed an accuracy of up to 95.5% with two optimal MSE scales (2D) and up to 97.7% with four optimal MSE scales (4D) in discriminating between young CHF and healthy participants. In older people (≥ 55 years), the discrimination accuracy reached 90.1% using LDA in 2D, SVM in 3D (three optimal MSE scales), and KNN in 5D (five optimal MSE scales). LDA with a 3D exhaustive search also achieved 94.4% accuracy in older people. Therefore, the results indicate that MSE analysis can differentiate between CHF and healthy individuals of any age.

Keywords: heart rate variability; multiscale entropy; heart failure; machine learning; low-dimensional exhaustive search; feature selection

1. Introduction

Heart disease is common and is associated with high mortality, as the end phase of many cardiovascular diseases is heart failure [1,2]. Heart failure in the elderly has a prevalence greater than 10% [3] and is a frequent cause of hospitalization [4]; the annual mortality rate of heart failure is 10% as well [5,6]. In addition, 30% to 40% of patients die within one year of being diagnosed with heart failure [7], and conditional four-year survivability is less than 50% [8]. Physiologically, heart failure is associated with dysfunction in the autonomic nervous system [9,10]. Autonomic activity is traditionally evaluated by heart rate variability (HRV) [11–13].

HRV is the physiological phenomenon of variation in the time interval between two consecutive heartbeats. The beat-to-beat interval is usually determined by the time interval between two consecutive R-peaks observed in an electrocardiogram (ECG), which is known as an RR interval. RR intervals between normal heartbeats, which excludes missed beat detections and premature beats, are referred to as NN intervals. RR interval and NN interval time series are two common heart rate signals for evaluating HRV.

HRV is analyzed from three perspectives—time domain, frequency domain, and nonlinear methods. All of these methods, however, characterize HRV based on a single temporal scale, which can result in a lack of data on features that can only be observed using alternate temporal scales.

Multiscale entropy (MSE) [14,15], developed from approximate entropy [16] and sample entropy [17], was thereby proposed as a method for measuring HRV across multiple temporal scales. Recently, many existing single-scale measures have been developed to multiscale measures, including multiscale fuzzy entropy [18–20], multiscale dispersion entropy [21–24], multiscale distribution entropy [25,26], multiscale permutation entropy [27–29], entropy of entropy, and average entropy [30–32]. Each specifically scaled measure reveals a particular feature of heartbeat signals. Measures at various scales representing multiple features form multidimensional HRV metrics. Machine learning is then employed to analyze multidimensional metrics and automatically discriminate between healthy people and those in heart failure, which has potential applications in diagnostic engines [33] for supporting physicians.

Machine learning is widely used in analyzing multidimensional HRV metrics to detect heart failure [34–42], and most studies combine several different HRV measures. However, these HRV metrics are not applied to clinically diagnose heart failure, which means that the improvement of HRV analysis is still required. Besides, the use of MSE for discrimination between congestive heart failure (CHF) and healthy hearts in older people remains controversial [43–46]. Relatively few studies have evaluated MSE with machine learning in the detection of heart failure. Costa and Healey [47] applied 40-dimensional MSE analysis and Fisher discriminant analysis to identify CHF, young healthy, and old healthy participants. Training accuracy reached 94% for the differentiation between CHF patients and old healthy participants. They also reported a positive predictivity of 76% and a specificity of 83%. In addition, they observed overtraining in using the 40 dimensions. Chen and Yang [48] used sequential forward feature selection to optimally choose a subset of 20 MSE scales, and reported sensitivities and specificities of 93.2 and 64.2% using linear discriminant analysis (LDA), 91.1 and 78.3% using quadratic discriminant analysis, and 95.6 and 64.7% using k-nearest neighbor (KNN), respectively. Liu and Gao [49] analyzed the RR interval time series and differential RR interval time series by MSE with scales ranging from 1 to 10 along with a support vector machine (SVM) classifier. Their reported accuracy was 75.9% for the analysis of the RR interval time series and 85.6% for the differential RR interval time series. Awan et al. [50] calculated MSE with scales from 1 to 20 and then utilized SVM, KNN, and random forest classifiers for discrimination, achieving testing accuracies of less than 85%. To the best of our knowledge, no study has reported a testing accuracy greater than 86% using MSE alone.

Certain studies directly applied several HRV indices for discrimination between healthy and CHF participants. However, excessive HRV indices in machine learning may lead to overtraining, which results in high training accuracy but low testing accuracy. To avoid overtraining, feature selection, a pre-process to reduce trivial dimensions, is executed before classification. The simplest feature selection is an exhaustive search that can also determine the optimal combination of dimensions. However, in high numbers of dimensions, the exhaustive search method becomes time-consuming and inaccessible. For example, conducting an exhaustive search for the optimal six dimensions (6D) among 20 leads to 38,760 dimensional combinations, meaning machine learning processing must be executed 38,760 times. As a result, the exhaustive search method is usually ignored, although an exhaustive search on just a few dimensions may be informative.

This study investigated MSE analysis accuracy with an exhaustive search of two, three, four, and five dimensions in differentiating younger (<55 years) and older (≥ 55 years) CHF patients from healthy ones in their respective age groups. LDA, SVM, and KNN classifiers were adopted for comparison with previous studies.

2. Materials and Methods

Figure 1 shows the flowchart of the analytical procedure used in this study. For discriminations between younger/older normal sinus rhythm (NSR) and congestive heart failure (CHF) participants, their preprocessed NN interval time series were analyzed by an MSE algorithm with r 0.10/0.15/0.20 and scales from 1 to 20 to extract 20 features. These 20 features were then evaluated by a low-dimensional exhaustive search to determine the optimal combinations of two, three, four, or five features. With respect to the 2D exhaustive search, for example, an unexamined combination of two features was

selected and the two features were used for discrimination using the LDA/SVM/KNN classifier with leave-one-out cross validation. This step was repeated to evaluate all combinations of two features. Afterward, combinations that led to overtraining were excluded, and then the optimal combinations of two features yielding the highest testing accuracy were determined from the rest.

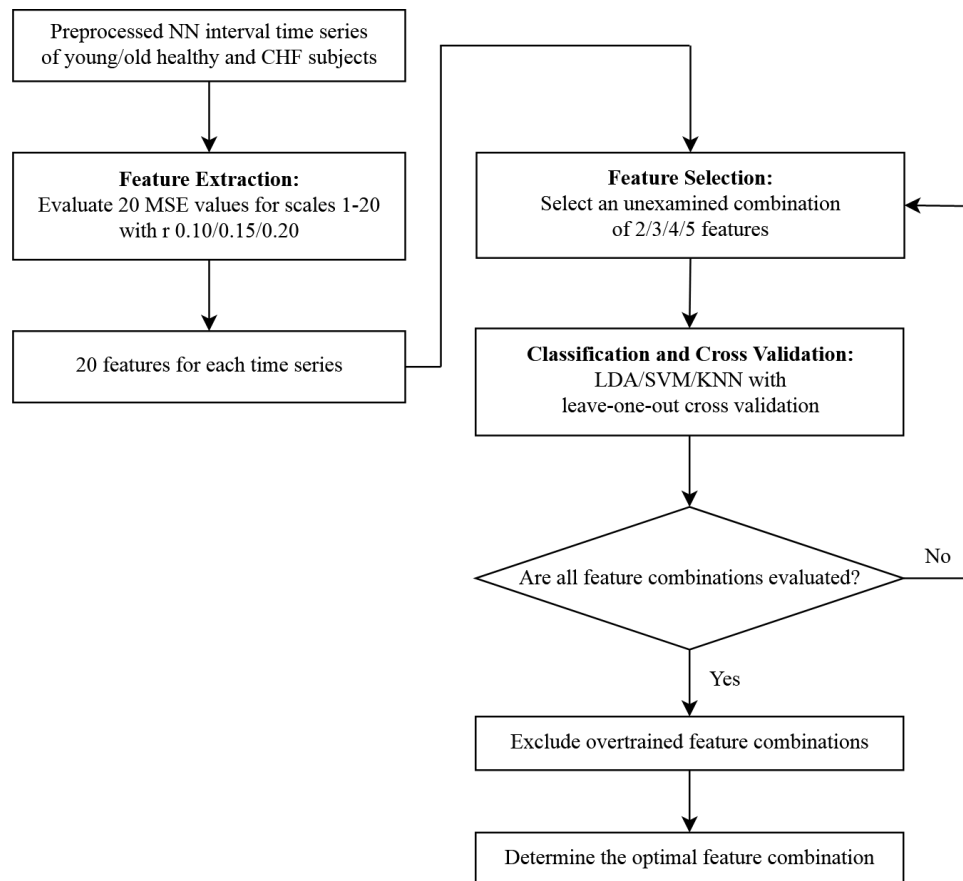


Figure 1. Flowchart of analytical procedure. Note: CHF, congestive heart failure; MSE, multiscale entropy; LDA, linear discriminant analysis; SVM, support vector machine; KNN, k-nearest neighbor.

2.1. Multiscale Entropy

MSE was used to extract features of heart rate signals (NN interval time series) in this study. MSE is the sample entropy from a coarse-grained time series [14,15]. Conceptually, sample entropy evaluates the difference between the number of repeating patterns of m length and the number of repeating patterns of $m + 1$ length [13]. The coarse-grain process resamples the original time series at a temporal scale τ in two steps. First, the original time series is equally divided into segments, each with a length of τ . Second, each segment is averaged, and these means form the coarse-grained time series at scale τ . In this study, MSE was calculated with an m of 2 [51] on scales ranging from 1 to 20.

To count the number of repeating patterns, the similarity criterion of two data points was defined as $r \times SD$, with SD being the standard deviation of the original time series. Similarity factor r usually varies from 0.10 to 0.20 [15] and has three suggested values, 0.10 [16,47,49,52], 0.15 [14–16,47,51,52], and 0.20 [25,53]; these three r values were thus employed in this study.

The number of data points in the time series for MSE analysis depends upon the value of m . To obtain a reliable MSE evaluation, this number of data points must be at least 10^m , and preferably at least 30^m [51]; for this study, 10,000, 20,000, and 40,000 data points were considered.

2.2. Data

The NN interval time series were taken from the following databases on PhysioNet [54]: nsrdb, MIT-BIH Normal Sinus Rhythm Database, five men aged 26 to 45 years and 13 women aged 20 to 50 years; nsr2db, Normal Sinus Rhythm RR Interval Database, 30 men aged 29 to 76 years and 24 women aged 58 to 73 years; chfdb, BIDMC Congestive Heart Failure Database, 11 men aged 22 to 71 years and four women aged 54 to 63 years; and chf2db, Congestive Heart Failure RR Interval Database, 29 patients aged 34 to 79 years including eight men, two women, and 19 patients of unknown gender.

These NN interval time series were categorized into four groups as follows: the young NSR group, 26 healthy participants younger than 55 years from nsrdb and nsr2db; the old NSR group, 46 healthy participants of ages equal to or older than 55 years from nsr2db; the young CHF group, 18 patients younger than 55 years from chfdb and chf2db; and the old CHF group, 25 patients equal with an age to or older than 55 years from chfdb and chf2db. One patient of unknown age in chfdb was excluded. All NN interval time series were preprocessed to remove artifacts [47]. Three sets of data points, numbering 10,000, 20,000, and 40,000, were used as described above.

2.3. Machine Learning

Three machine learning classifiers, LDA, SVM, and KNN, were applied to automatically discriminate between young CHF and healthy participants, and old CHF and healthy participants. In the SVM algorithm, the radial basis function (rbf) kernel was used. For the KNN classifier, the optimized number of neighbors was three in this study. The inputs for the three classifiers were the two, three, four, and five features out of 20 of the preprocessed NN interval time series, extracted by the MSE algorithm for 2D, 3D, 4D, and 5D exhaustive searches, respectively.

The optimal MSE scales were determined for each classifier by means of exhaustive search in two, three, four, or five dimensions (denoted as 2D, 3D, 4D, and 5D, respectively) among 20 (20 MSE scales). Optimization was examined via leave-one-out cross validation, which is suited for small sample sizes. To reduce overtraining effects, an overtraining tolerance of 5% was chosen, defined as the absolute difference between the training and testing accuracies.

2.4. Performance Metrics

The performance of MSE analysis with machine learning with few dimensions was evaluated using five traditional indices: accuracy (Acc), sensitivity (Sen), specificity (Spe), positive predictive value (PPV), and negative predictive value (NPV). According to [55,56], the correct classification rate for class k , CCR_k , is defined as:

$$CCR_k = \frac{T_k}{N_k} \times 100\%, \quad (1)$$

where T_k is the number of correct identifications of class k and N_k is the total number of participants of class k . Thus, Acc, Sen, and Spe can be separately defined by:

$$Acc = \frac{\sum_k CCR_k \cdot N_k}{\sum_k N_k} = \frac{T_{NSR} + T_{CHF}}{N_{NSR} + N_{CHF}} \times 100\%, \quad (2)$$

$$Sen = CCR_{CHF} = \frac{T_{CHF}}{N_{CHF}} \times 100\%, \quad (3)$$

$$Spe = CCR_{NSR} = \frac{T_{NSR}}{N_{NSR}} \times 100\%. \quad (4)$$

PPV and NPV are separately determined by:

$$PPV = \frac{T_{CHF}}{T_{CHF} + (N_{NSR} - T_{NSR})} \times 100\%, \quad (5)$$

$$NPV = \frac{T_{NSR}}{T_{NSR} + (N_{CHF} - T_{CHF})} \times 100\%. \tag{6}$$

3. Results

3.1. Multiscale Entropy Evaluation

Figure 2 shows that the average MSE values at different scales in the NN interval time series of young NSR participants (black circles) deviated from those of the other three groups (old NSR, young CHF, and old CHF participants). This suggests that discrimination between NSR and CHF is relatively easy in younger individuals (<55 years), and becomes more difficult as they age (≥55 years). For a fixed r value and scale τ , MSE values with different data sample sizes are similar to one another. Conversely, for the same data sample size and scale τ , MSE values decrease as r increases.

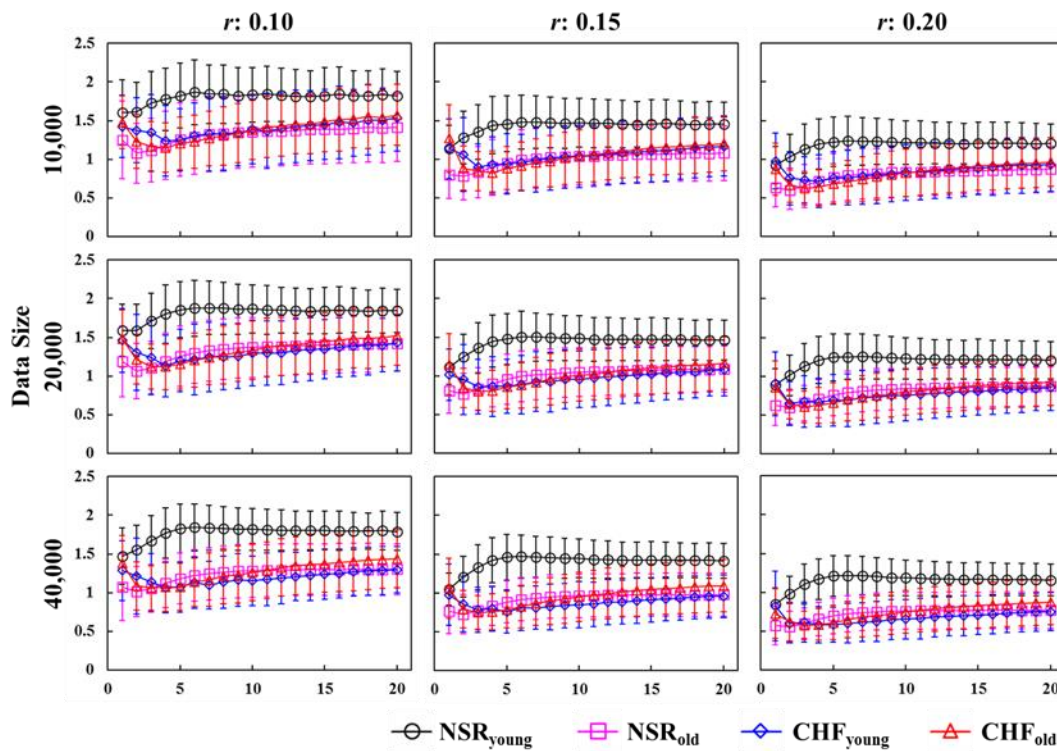


Figure 2. Multiscale entropy analysis (scales from 1 to 20) of NN interval time series from the young normal sinus rhythm (NSR) group (NSR_{young}; black circles), old NSR group (NSR_{old}; purple squares), young congestive heart failure (CHF) group (CHF_{young}; blue diamonds), and old CHF group (CHF_{old}; red triangles). Error bars represent standard deviations.

3.2. Discrimination between Younger NSR and CHF Participants

Table 1 shows the discrimination accuracies between young NSR and CHF participants using LDA, SVM, and KNN. Among 20 MSE scales, the optimal combination of two, three, four, or five MSE scales was determined by means of an exhaustive search. LDA appeared to be superior to SVM and KNN, with an accuracy of 95.5% with only two MSE scales (2D), an r of 0.10, and a data sample size of 40,000. For the same data sample size and classifier, accuracy reached 97.7% with four MSE scales (4D) and an r of 0.10 or 0.20.

Table 1. Discrimination accuracy between young NSR and CHF participants using linear discriminant analysis (LDA), support vector machine (SVM), and k-nearest neighbor (KNN). Note: SD, standard deviation; 2D, two optimal MSE scales; 3D, three optimal MSE scales; 4D, four optimal MSE scales; 5D, five optimal MSE scales.

| LDA | | | | | | | | | | | | |
|-----------|------------|------|------|------------|------|------|------------|------|-------------------|------------|------|------|
| <i>r</i> | 2D | | | 3D | | | 4D | | | 5D | | |
| | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 |
| 40,000 | 95.5 | 90.9 | 93.2 | 95.5 | 95.5 | 95.5 | 97.7 | 95.5 | 97.7 ¹ | 97.7 | 97.7 | 97.7 |
| 20,000 | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | 88.6 | 88.6 | 88.6 | 90.9 ¹ | 88.6 | 88.6 | 90.9 |
| 10,000 | 86.4 | 86.4 | 89.6 | 88.6 | 90.9 | 90.9 | 88.6 | 93.2 | 93.2 ¹ | 90.9 | 93.2 | 93.2 |
| Mean ± SD | 83.6 ± 4.7 | | | 90.9 ± 3.6 | | | 92.7 ± 3.5 | | | 93.2 ± 3.6 | | |
| SVM | | | | | | | | | | | | |
| <i>r</i> | 2D | | | 3D | | | 4D | | | 5D | | |
| | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 |
| 40,000 | 93.2 | 90.9 | 90.9 | 93.2 | 90.9 | 90.9 | 90.9 | 90.9 | 90.9 | 90.9 | 90.9 | 90.9 |
| 20,000 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 |
| 10,000 | 81.8 | 79.5 | 84.1 | 86.4 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 |
| Mean ± SD | 87.4 ± 4.3 | | | 89.1 ± 2.1 | | | 88.6 ± 1.9 | | | 88.6 ± 1.8 | | |
| KNN | | | | | | | | | | | | |
| <i>r</i> | 2D | | | 3D | | | 4D | | | 5D | | |
| | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 |
| 40,000 | 88.6 | 88.6 | 90.9 | 90.9 | 88.6 | 90.9 | 90.9 | 88.6 | 90.9 | 88.6 | 88.6 | 90.9 |
| 20,000 | 86.4 | 86.4 | 86.4 | 86.4 | 84.1 | 88.6 | 88.6 | 86.4 | 86.4 | 88.6 | 86.4 | 86.4 |
| 10,000 | 81.8 | 84.1 | 84.1 | 81.8 | 84.1 | 86.4 | 84.1 | 86.4 | 86.4 | 88.6 | 86.4 | 86.4 |
| Mean ± SD | 86.4 ± 2.6 | | | 86.9 ± 3.0 | | | 87.6 ± 2.2 | | | 87.9 ± 1.5 | | |

¹ Highest accuracies are shown for the least number of dimensions using the same *r* value (0.20) and machine learning method (LDA).

For data sample sizes of 40,000, 20,000, and 10,000, the highest accuracies were 97.7, 90.9, and 93.2%, respectively. These accuracies were also achieved using an *r* of 0.2 and 4D exhaustive search with the LDA classifier, as indicated in the footnotes of Table 1. The 4D combinations are listed in Table 2, along with their Acc, Sen, Spe, PPV, and NPV. The combination of MSE scales, {4, 5, 6, 9}, yielded 100% Spe and PPV for a data sample size 40,000.

Table 2. List of the combinations of the optimal four MSE (*r* = 0.20) scales for discrimination between young NSR and CHF participants using LDA, as well as their corresponding accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. Note: Acc, accuracy; Sen, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value.

| Data Sample Sizes | Scales | Acc | Sen | Spe | PPV | NPV |
|-------------------|----------------|------|------|-------|-------|------|
| 40,000 | {4, 5, 6, 9} | 97.7 | 94.4 | 100.0 | 100.0 | 96.3 |
| 20,000 | {1, 6, 9, 13} | 90.9 | 83.3 | 96.2 | 93.8 | 89.3 |
| | {1, 6, 9, 18} | 90.9 | 83.3 | 96.2 | 93.8 | 89.3 |
| | {1, 6, 9, 19} | 90.9 | 83.3 | 96.2 | 93.8 | 89.3 |
| 10,000 | {3, 4, 11, 16} | 93.2 | 88.9 | 96.2 | 94.1 | 92.6 |

3.3. Discrimination between Older NSR and CHF Participants

Table 3 shows the discrimination accuracies between NSR and CHF in older individuals under various conditions. Even though average MSE values from the NN interval time series for older NSR and CHF participants did not appear different from one another (Figure 2), accuracy reached 90.1%

using LDA in 2D, SVM in 3D, and KNN in 5D. Overtraining occurred for eight conditions, labeled NA in Table 3, and the results showed marked overtraining for KNN. This finding suggests that the KNN classifier combined with MSE analysis is not appropriate for differentiating between older CHF and NSR participants with fewer dimensions.

Table 3. Discrimination accuracy between older NSR and CHF participants using LDA, SVM, and KNN.

| LDA | | | | | | | | | | | | |
|-----------|-----------------|------|------|------------|-------------------|------|------------|------|------|------------|------|------|
| <i>r</i> | 2D | | | 3D | | | 4D | | | 5D | | |
| | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 |
| 40,000 | 88.7 | 88.7 | 90.1 | 90.1 | 94.4 ² | 91.5 | 91.5 | 93.0 | 93.0 | 91.5 | 93.0 | 93.0 |
| 20,000 | 83.1 | 84.5 | 84.5 | 87.3 | 90.1 ² | 88.7 | 88.7 | 90.1 | 87.3 | 88.7 | 90.1 | 88.7 |
| 10,000 | 83.1 | 84.5 | 84.5 | 90.1 | 93.0 ² | 90.1 | 88.7 | 93.0 | 93.0 | 90.1 | 93.0 | 93.0 |
| Mean ± SD | 85.8 ± 2.5 | | | 90.6 ± 2.0 | | | 90.9 ± 2.1 | | | 91.2 ± 1.7 | | |
| SVM | | | | | | | | | | | | |
| <i>r</i> | 2D | | | 3D | | | 4D | | | 5D | | |
| | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 |
| 40,000 | 85.9 | 87.3 | 85.9 | 90.1 | 88.7 | 87.3 | 90.1 | 90.1 | 87.3 | 91.5 | 90.1 | 87.3 |
| 20,000 | 80.3 | 81.7 | 81.7 | 83.1 | 88.7 | 83.1 | 85.9 | 88.7 | 83.1 | 87.3 | 88.7 | 85.9 |
| 10,000 | 80.3 | 81.7 | 78.9 | 87.3 | 90.1 | 88.7 | 87.3 | 91.5 | 91.5 | 88.7 | 91.5 | 90.1 |
| Mean ± SD | 82.6 ± 2.8 | | | 87.5 ± 2.5 | | | 88.4 ± 2.6 | | | 89.0 ± 1.8 | | |
| KNN | | | | | | | | | | | | |
| <i>r</i> | 2D | | | 3D | | | 4D | | | 5D | | |
| | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 | 0.10 | 0.15 | 0.20 |
| 40,000 | NA ³ | NA | NA | 84.5 | 84.5 | 85.9 | 85.9 | 88.7 | 85.9 | 85.9 | 85.9 | 85.9 |
| 20,000 | 77.5 | NA | NA | 81.7 | NA | NA | 83.1 | 76.1 | 81.7 | 84.5 | NA | 83.1 |
| 10,000 | 77.5 | 78.9 | 85.9 | 84.5 | 84.5 | 84.5 | 83.1 | 87.3 | 88.7 | 84.5 | 88.7 | 90.1 |
| Mean ± SD | 79.9 ± 3.5 | | | 84.3 ± 1.2 | | | 84.5 ± 3.8 | | | 86.1 ± 2.1 | | |

² Highest accuracies for the least number of dimensions were found for the same *r* (0.15) and machine learning method (LDA) among different data sample sizes. ³ NA: instances where overtraining occurred for all combinations of dimensions.

On average, LDA showed better accuracy than SVM or KNN. LDA accuracies were 94.4, 90.1, and 93.0% for data sample sizes of 40,000, 20,000, and 10,000, respectively. The utilization of an *r* of 0.15 and a 3D exhaustive search also yielded the highest accuracies. Combinations of the three MSE scales (3D) and their performances are listed in Table 4. Moreover, the combinations of {3, 6, 12} for 40,000 points, {2, 6, 9} for 20,000 points, and {1, 9, 18} for 10,000 points were associated with 100% Spe and PPV.

Table 4. List of combinations of three optimal MSE (*r* = 0.15) scales for the discrimination between older NSR and CHF participants using LDA, as well as the corresponding accuracy, sensitivity, specificity, positive predictive value, and negative predictive value.

| Data Sample Sizes | Scales | Acc | Sen | Spe | PPV | NPV |
|-------------------|------------|------|------|-------|-------|------|
| 40,000 | {3, 6, 12} | 94.4 | 84.0 | 100.0 | 100.0 | 92.0 |
| 20,000 | {1, 6, 12} | 90.1 | 76.0 | 97.8 | 95.0 | 88.2 |
| | {1, 6, 13} | 90.1 | 76.0 | 97.8 | 95.0 | 88.2 |
| | {1, 6, 18} | 90.1 | 76.0 | 97.8 | 95.0 | 88.2 |
| | {2, 6, 9} | 90.1 | 72.0 | 100.0 | 100.0 | 86.8 |
| 10,000 | {1, 9, 18} | 93.0 | 80.0 | 100.0 | 100.0 | 90.2 |

4. Discussion

Applying MSE analysis to differentiate between older CHF and healthy participants has been controversial. Some reports have indicated that the five and six MSE scales are significantly different between the CHF and NSR groups [47,49]. Others have reported that older CHF and NSR participants cannot be differentiated by MSE alone [43–46]. Our study shows that the {6} MSE scale is important because it is frequently involved, as shown in Tables 2 and 4; however, a single MSE scale is not sufficient for discrimination. This study indicates that a combination of three or four MSE scales (Tables 2 and 4) is preferred.

Regarding MSE analysis, one difficulty is the pre-determination of the similarity factor r . Three r values—0.10 [16,47,49,52], 0.15 [14–16,47,51,52], and 0.20 [25,53]—have been widely used. The results of this study suggest that age is critical to r . According to Tables 1 and 3, we suggest values of 0.15 and 0.20 for discrimination in older and younger participants, respectively. This should be confirmed with larger datasets in the future.

Many methods have been used for feature selection in preprocessing multidimensional HRV metrics, such as sequential forward feature selection [48,57], sequential backward feature selection [42], and genetic algorithms [39]. Up until now, the use of an exhaustive search has been ignored because it is extremely time-consuming in higher dimensions. We simplified it and used it for low-dimensional exhaustive searches, where it demonstrated excellent utility in MSE analysis for the discrimination between older CHF and healthy participants.

5. Conclusions

Machine learning has been widely used in analyzing multidimensional HRV metrics to detect heart failure. The use of MSE for the discrimination between heart failure and healthy hearts in older people remains controversial. No study has reported a testing accuracy greater than 86% using MSE alone. This study illustrates that a low-dimensional exhaustive search improves the determination between CHF and NSR individuals using MSE analysis. Accuracies were found to be greater than 90% within 5% overtraining tolerance. For discrimination in younger individuals (<55 years), MSE analysis with an r of 0.20 combined with LDA in a 4D exhaustive search is suggested. This yielded accuracies of 97.7, 90.9, and 93.2% for data sample sizes of 40,000, 20,000, and 10,000, respectively. For discrimination in older participants (≥ 55 years), MSE analysis with an r of 0.15 along with LDA in a 3D exhaustive search is suggested. These accuracies were 94.4, 90.1, and 93.0% for data sample sizes of 40,000, 20,000, and 10,000, respectively. The results indicate that MSE analysis could extract important features hidden in the heart rate signals of healthy and heart failure participants. Three or four features are also sufficient for discrimination, with a testing accuracy greater than 90%. Therefore, MSE analysis combined with machine learning and a low-dimensional exhaustive search can differentiate between CHF and healthy individuals of any age.

Practically, cardiologists do not use HRV metrics to diagnose CHF at present. Previous research and this work aim to improve HRV analysis for clinical application. With respect to MSE analysis, this work uses the low-dimensional exhaustive search to determine the optimal three and four features, which yielded the highest testing accuracy compared to previous studies. The analysis algorithm of this work and the optimal features can be applied to diagnostic engines for further clinical research and supporting cardiologists.

Author Contributions: Conceptualization, H.-H.C. and C.-W.Y.; methodology, H.-H.C.; formal analysis, H.-H.C., C.-W.Y. and C.F.H.; writing—original draft preparation, H.-H.C.; writing—review and editing, L.H. and S.C.; supervision, L.H. and S.C.

Funding: This research was funded by the Ministry of Science and Technology of the Republic of China, grant number MOST107-2221-E009-016.

Acknowledgments: We appreciate the three reviewers and the academic editor whose valuable comments led to a much improved version of this paper.

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

References

1. McMurray, J.J.; Pfeffer, M.A. Heart failure. *Lancet* **2005**, *365*, 1877–1889. [[CrossRef](#)]
2. National Clinical Guideline Centre (UK). *Chronic Heart Failure: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care*; Royal College of Physicians: London, UK, 2010.
3. Ponikowski, P.; Voors, A.A.; Anker, S.D.; Bueno, H.; Cleland, J.G.; Coats, A.J.; Falk, V.; Gonzalez-Juanatey, J.R.; Harjola, V.P.; Jankowska, E.A.; et al. 2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure: The task force for the diagnosis and treatment of acute and chronic heart failure of the european society of cardiology (esc). Developed with the special contribution of the heart failure association (hfa) of the esc. *Eur. J. Heart Fail.* **2016**, *18*, 891–975. [[PubMed](#)]
4. Gottdiener, J.S.; McClelland, R.L.; Marshall, R.; Shemanski, L.; Furberg, C.D.; Kitzman, D.W.; Cushman, M.; Polak, J.; Gardin, J.M.; Gersh, B.J.; et al. Outcome of congestive heart failure in elderly persons: Influence of left ventricular systolic function—The cardiovascular health study. *Ann. Int. Med.* **2002**, *137*, 631–639. [[CrossRef](#)] [[PubMed](#)]
5. Cleland, J.G.F.; Daubert, J.; Erdmann, E.; Freemantle, N.; Gras, D.; Kappenberger, L.; Tavazzi, L.; Cleland, J.G.F.; Daubert, J.C.; Erdmann, E.; et al. The effect of cardiac resynchronization on morbidity and mortality in heart failure. *N. Engl. J. Med.* **2005**, *352*, 1539–1549. [[CrossRef](#)] [[PubMed](#)]
6. Nolan, J.; Batin, P.D.; Andrews, R.; Lindsay, S.J.; Brooksby, P.; Mullen, H.; Baig, W.; Flapan, A.D.; Cowley, A.; Prescott, R.J.; 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). *Circulation* **1998**, *98*, 1510–1516. [[CrossRef](#)] [[PubMed](#)]
7. Neubauer, S. Mechanisms of disease—The failing heart—An engine out of fuel. *N. Engl. J. Med.* **2007**, *356*, 1140–1151. [[CrossRef](#)]
8. Lane, R.E.; Cowie, M.R.; Chow, A.W.C. Prediction and prevention of sudden cardiac death in heart failure. *Heart* **2005**, *91*, 674–680. [[CrossRef](#)]
9. Kishi, T. Heart failure as an autonomic nervous system dysfunction. *J. Cardiol.* **2012**, *59*, 117–122. [[CrossRef](#)]
10. Triposkiadis, F.; Karayannis, G.; Giamouzis, G.; Skoularigis, J.; Louridas, G.; Butler, J. The sympathetic nervous system in heart failure physiology, pathophysiology, and clinical implications. *J. Am. Coll. Cardiol.* **2009**, *54*, 1747–1762. [[CrossRef](#)]
11. Camm, A.J.; Malik, M.; Bigger, J.T.; Breithardt, G.; Cerutti, S.; Cohen, R.J.; Coumel, P.; Fallen, E.L.; Kennedy, H.L.; Kleiger, R.E.; et al. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur. Heart J.* **1996**, *17*, 354–381.
12. Pumplrla, J.; Howorka, K.; Groves, D.; Chester, M.; Nolan, J. Functional assessment of heart rate variability: Physiological basis and practical applications. *Int. J. Cardiol.* **2002**, *84*, 1–14. [[CrossRef](#)]
13. Gutierrez-Tobal, G.C.; Alvarez, D.; Gomez-Pilar, J.; del Campo, F.; Hornero, R. Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women. *Entropy* **2015**, *17*, 123–141. [[CrossRef](#)]
14. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [[CrossRef](#)] [[PubMed](#)]
15. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906. [[CrossRef](#)] [[PubMed](#)]
16. Pincus, S.M. Approximate entropy as a measure of system-complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)] [[PubMed](#)]
17. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circ.* **2000**, *278*, H2039–H2049. [[CrossRef](#)] [[PubMed](#)]
18. Chen, W.T.; Zhuang, J.; Yu, W.X.; Wang, Z.Z. Measuring complexity using fuzzyen, apen, and sampen. *Med. Eng. Phys.* **2009**, *31*, 61–68. [[CrossRef](#)] [[PubMed](#)]

19. Zheng, J.D.; Cheng, J.S.; Yang, Y.; Luo, S.R. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mech. Mach. Theory* **2014**, *78*, 187–200. [[CrossRef](#)]
20. Li, M.A.; Liu, H.N.; Zhu, W.; Yang, J.F. Applying improved multiscale fuzzy entropy for feature extraction of mi-eeg. *Appl. Sci.* **2017**, *7*, 92. [[CrossRef](#)]
21. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Proc. Lett.* **2016**, *23*, 610–614. [[CrossRef](#)]
22. Azami, H.; Rostaghi, M.; Abasolo, D.; Escudero, J. Refined composite multiscale dispersion entropy and its application to biomedical signals. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2872–2879.
23. Zhang, Y.D.; Tong, S.G.; Cong, F.Y.; Xu, J. Research of feature extraction method based on sparse reconstruction and multiscale dispersion entropy. *Appl. Sci.* **2018**, *8*, 888. [[CrossRef](#)]
24. Fazan, F.S.; Brognara, F.; Fazan, R.; Murta, L.O.; Silva, L.E.V. Changes in the complexity of heart rate variability with exercise training measured by multiscale entropy-based measurements. *Entropy* **2018**, *20*, 47. [[CrossRef](#)]
25. Li, P.; Liu, C.Y.; Li, K.; Zheng, D.C.; Liu, C.C.; Hou, Y.L. Assessing the complexity of short-term heartbeat interval series by distribution entropy. *Med. Biol. Eng. Comput.* **2015**, *53*, 77–87. [[CrossRef](#)]
26. Lee, D.Y.; Choi, Y.S. Multiscale distribution entropy analysis of short-term heart rate variability. *Entropy* **2018**, *20*, 952. [[CrossRef](#)]
27. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)]
28. Ouyang, G.X.; Li, J.; Liu, X.Z.; Li, X.L. Dynamic characteristics of absence eeg recordings with multiscale permutation entropy analysis. *Epilepsy Res.* **2013**, *104*, 246–252. [[CrossRef](#)]
29. Gao, Y.D.; Vilecco, F.; Li, M.; Song, W.Q. Multi-scale permutation entropy based on improved lmd and hmm for rolling bearing diagnosis. *Entropy* **2017**, *19*, 176. [[CrossRef](#)]
30. Shannon, C.E. The mathematical theory of communication (reprinted). *MD Comput.* **1997**, *14*, 306–317.
31. Hsu, C.F.; Wei, S.Y.; Huang, H.P.; Hsu, L.; Chi, S.; Peng, C.K. Entropy of entropy: Measurement of dynamical complexity for biological systems. *Entropy* **2017**, *19*, 550. [[CrossRef](#)]
32. Hsu, C.F.; Lin, P.Y.; Chao, H.H.; Hsu, L.; Chi, S. Average entropy: Measurement of disorder for cardiac rr interval signals. *Physica A* **2019**, *529*, 1215333. [[CrossRef](#)]
33. Sena, P.; Attianese, P.; Pappalardo, M.; Vilecco, F. FIDELITY: Fuzzy Inferential Diagnostic Engine for on-Line Support to Physicians. In Proceedings of the 4th International Conference on Biomedical Engineering in Vietnam, Ho Chi Minh City, Vietnam, 8–10 January 2012; Toi, V., Toan, N., Dang Khoa, T., Lien Phuong, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 396–400.
34. Acharya, U.R.; Fujita, H.; Sudarshan, V.K.; Oh, S.L.; Muhammad, A.; Koh, J.E.W.; Tan, J.H.; Chua, C.K.; Chua, K.P.; Tan, R.S. Application of empirical mode decomposition (emd) for automated identification of congestive heart failure using heart rate signals. *Neural Comput. Appl.* **2017**, *28*, 3073–3094. [[CrossRef](#)]
35. Melillo, P.; Fusco, R.; Sansone, M.; Bracale, M.; Pecchia, L. Discrimination power of long-term heart rate variability measures for chronic heart failure detection. *Med. Biol. Eng. Comput.* **2011**, *49*, 67–74. [[CrossRef](#)] [[PubMed](#)]
36. Pecchia, L.; Melillo, P.; Sansone, M.; Bracale, M. Discrimination power of short-term heart rate variability measures for chf assessment. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 40–46. [[CrossRef](#)]
37. Pecchia, L.; Melillo, P.; Bracale, M. Remote health monitoring of heart failure with data mining via cart method on hrv features. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 800–804. [[CrossRef](#)]
38. Kamath, C. A new approach to detect congestive heart failure using teager energy nonlinear scatter plot of r-r interval series. *Med. Eng. Phys.* **2012**, *34*, 841–848. [[CrossRef](#)] [[PubMed](#)]
39. Yu, S.N.; Lee, M.Y. Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. *Comput. Biol. Med.* **2012**, *42*, 816–825. [[CrossRef](#)] [[PubMed](#)]
40. Melillo, P.; De Luca, N.; Bracale, M.; Pecchia, L. Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. *IEEE J. Biomed. Health* **2013**, *17*, 727–733. [[CrossRef](#)]
41. Liu, G.Z.; Wang, L.; Wang, Q.; Zhou, G.M.; Wang, Y.; Jiang, Q. A new approach to detect congestive heart failure using short-term heart rate variability measures. *PLoS ONE* **2014**, *9*, e93399. [[CrossRef](#)] [[PubMed](#)]

42. Narin, A.; Isler, Y.; Ozer, M. Investigating the performance improvement of HRV indices in CHF using feature selection methods based on backward elimination and statistical significance. *Comput. Biol. Med.* **2014**, *45*, 72–79. [[CrossRef](#)] [[PubMed](#)]
43. Kuntamalla, S.; Lekkala, R.G.R. Reduced data dualscale entropy analysis of HRV signals for improved congestive heart failure detection. *Meas. Sci. Rev.* **2014**, *14*, 294–301. [[CrossRef](#)]
44. Huang, X.L.; Ning, X.B.; Wang, X.L. Multiscale analysis of heart beat interval increment series and its clinical significance. *Chin. Sci. Bull.* **2009**, *54*, 3784–3789. [[CrossRef](#)]
45. Wessel, N.; Schirdewan, A.; Kurths, J. Intermittently decreased beat-to-beat variability in congestive heart failure. *Phys. Rev. Lett.* **2003**, *91*, 119801. [[CrossRef](#)] [[PubMed](#)]
46. Thuraisingham, R.A.; Gottwald, G.A. On multiscale entropy analysis for physiological data. *Physica A* **2006**, *366*, 323–332. [[CrossRef](#)]
47. Costa, M.; Healey, J.A. Multiscale entropy analysis of complex heart rate dynamics: Discrimination of age and heart failure effects. *Comput. Cardiol.* **2003**, *30*, 705–708.
48. Chen, Y.; Yang, H. A comparative analysis of alternative approaches for quantifying nonlinear dynamics in cardiovascular system. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 2599–2602.
49. Liu, C.Y.; Gao, R. Multiscale entropy analysis of the differential rr interval time series signal and its application in detecting congestive heart failure. *Entropy* **2017**, *19*, 251.
50. Awan, I.; Aziz, W.; Shah, I.H.; Habib, N.; Alowibdi, J.S.; Saeed, S.; Nadeem, M.S.A.; Shah, S.A.A. Studying the dynamics of interbeat interval time series of healthy and congestive heart failure subjects using scale based symbolic entropy analysis. *PLoS ONE* **2018**, *13*, e0196823. [[CrossRef](#)]
51. Pincus, S.M.; Goldberger, A.L. Physiological time-series analysis—What does regularity quantify. *Am. J. Physiol.* **1994**, *266*, H1643–H1656. [[CrossRef](#)]
52. Zhao, L.N.; Wei, S.S.; Zhang, C.Q.; Zhang, Y.T.; Jiang, X.G.; Liu, F.; Liu, C.Y. Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects. *Entropy* **2015**, *17*, 6270–6288. [[CrossRef](#)]
53. Lake, D.E.; Richman, J.S.; Griffin, M.P.; Moorman, J.R. Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2002**, *283*, R789–R797. [[CrossRef](#)]
54. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. Physiobank, physiotoolkit, and physionet—Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, E215–E220. [[CrossRef](#)]
55. Cascio, D.; Taormina, V.; Cipolla, M.; Bruno, S.; Fauci, F.; Raso, G. A multi-process system for HEp-2 cells classification based on SVM. *Pattern Recognit. Lett.* **2016**, *82*, 56–63. [[CrossRef](#)]
56. Benammar Elgaaied, A.; Cascio, D.; Bruno, S.; Ciaccio, M.C.; Cipolla, M.; Fauci, A.; Morgante, R.; Taormina, V.; Gorgi, Y.; Marrakchi Triki, R.; et al. Computer-assisted classification patterns in autoimmune diagnostics: The AIDA Project. *BioMed Res. Int.* **2016**, *2016*, 2073076. [[CrossRef](#)]
57. Hua, Z.; Chen, C.; Zhang, R.; Liu, G.; Wen, W. Diagnosing various severity levels of congestive heart failure based on long-term HRV signal. *Appl. Sci.* **2019**, *9*, 2544.

