

Open Access

Faster Detection of Abnormal Electrocardiogram (ECG) Signals Using Fewer Features of Heart Rate Variability (HRV)

Gong X¹, Long B¹, Wang Z¹, Zhang H¹ and Nandi AK^{1,2*}

¹College of Electronic and Information Engineering, Tongji University, Shanghai, China

²Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, Middlesex, UK

Abstract

To reduce the effect of noise in raw Electrocardiogram (ECG) data for faster detection of cardiac arrhythmia, Heart Rate Variability (HRV) features represent good choices. This work extracted 34 popular Heart Rate Variability (HRV) features based on the MIT-BIH Arrhythmia Database. Combinations of 11 feature selection algorithms and 2 classification algorithms are used to discover the effective features of the abnormal ECG signal detection. The systematic comparisons show that the combination of 34 original features has a stable classification performance for 3 different time windows, i.e., 32 RR-intervals, 5 minutes, and 30 minutes of raw ECG records. It has been discovered that a 10-feature combination (RMSSD, SDNN, CV, TINN, HF, SampEn, SD1/SD2, VAI, ED, and DC) can rapidly classify the arrhythmia and normal state, based on the shortest ECG records (32 RR-intervals). The future work will utilize this combination of features to implement in a portable ECG equipment and clinical Arrhythmia on-line detection.

Keywords: Heart rate variability; Electrocardiogram; Feature selection; Support vector machine; Back propagation neural network

Introduction

Cardiovascular disease (CVD) has the highest mortality. According to the World Health Organization (WHO), approximately 17.7 million people died of cardiovascular disease in 2015 [1]. Automatic arrhythmia detection and classification can be achieved by many kinds of features, such as morphological features [2], a 2D representation of electrograms known as the Spatial Projection of Tachycardia (SPOT) [3], etc. Different kind of machine learning algorithms were used to achieved better detections performance of arrhythmia based on ECG morphological features, such as wavelet transform [4], fuzzy C-means (FCM) clustering [5], independent component analysis (ICA) [6], Support vector machines [7-10]. Moaveian et al. performed a qualitative comparison of a neural network and SVM for classifying ECG arrhythmia based on the performance and training time for beats [11], and they found that Support Vector Machine (SVM) was faster but a neural network was more stable. Arrhythmia classification based on morphological features is a time-consuming process, and the results are very sensitive to the amount of the noise. Heart rate variability (HRV) [12] is a slight fluctuation of the instantaneous heart rate between consecutive heartbeats. Those extracting features based on an HRV signal from the ECG recordings can overcome the influence of noise in the original ECG data. The combination of the linear and nonlinear features appeared to be a better choice to detect arrhythmia [13]. Deceleration capacity (DC) of the heart rate was also a good predictor of arrhythmia or even sudden cardiac death [14,15].

Recent developments in electronic technology and algorithms have provided the possibility to offer monitoring and early warning of cardiovascular disease using portable ECG equipment. Martis et al. reviewed the current methods (linear and non-linear features) in ECG which could be utilized on the portable ECG equipment [16]. Deep convolutional neural network was invited into automated Myocardial Infarction detection [17]. He et al. developed the lightweight neural network for detecting arrhythmias which might be used on mobile devices [18]. To detect arrhythmia with good features easily, effectively, and fast, remains a challenging mission. This work utilizes the famous MIT-BIH database [19], which has been used to detect automatically arrhythmia in many studies [7,18]. Overall, 34 features, including both linear and non-linear features as well as the DC of the heart rate, are extracted. A systematic performance comparison of combinations of 11 feature selection and 2 classification algorithms is carried out in this work to find the most effective HRV features. We investigate the influence of feature selection on the training and testing performance as well as the computation time for the SVM and the Back Propagation Neural Network (BPN) for detecting arrhythmia. Finally, we discover that a 10-feature combination that may suitable for a portable ECG device, which have a rapid computational speed and a high prediction accuracy with a short data acquisition time.

Materials and Methods

Database

The MIT-BIH Arrhythmia database was collected from 25 males aged between 32 and 89 years as well as 22 females aged between 23 and 89 years. This database has a total of 48 ECG records, with two records obtained from each subject. Each recording contains ECG data with a duration of more than 30 minutes. The MIT-BIH Normal sinus rhythm database was collected from 5 males aged between 26 and 45 years as well as 13 females aged between 20 and 50 years, for a total of 18 ECG recordings, each containing ECG data with a length of 24 hours. To ensure that the sizes of the data sets were consistent, we selected 1.5 hours of ECG data for each normal subject and segmented the data

*Corresponding author: Nandi AK, FREng, FIEEE, PhD (Cambridge), IEEE Distinguished Lecturer (EMBS, 2018-2019), Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, Middlesex, UK, Tel: +44(0)1895266119; Fax: +44(0)1895269782; E-mail: asoke.nandi@brunel.ac.uk

Received December 04, 2018; Accepted December 27, 2018; Published January 03, 2019

Citation: Gong X, Long B, Wang Z, Zhang H, Nandi AK (2018) Faster Detection of Abnormal Electrocardiogram (ECG) Signals Using Fewer Features of Heart Rate Variability (HRV). J Comput Sci Syst Biol 12: 19-27. doi:10.4172/jcsb.1000295

Copyright: © 2018 Gong X, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

J Comput Sci Syst Biol, an open access journal ISSN: 0974-7231

sets into 54 30-minute recordings, from which we finally acquired 48 recordings for further analysis.

To evaluate the classification performance on the short time data set, each 30-minute ECG recording was segmented into 5-minute pieces, for approximately 288 samples. In addition, the 5-minute recordings were segmented into 32 RR intervals, with approximately 2000 samples for both the arrhythmia and normal groups.

The proposed algorithms

The scheme of the proposed procedure is illustrated in Figure 1. The process has 5 steps: Pre-processing, Feature extraction, Normalization, Feature selection, and Arrhythmia classification. Pre-processing is directly carried out on PhysioNet [20] using the open-source software PhysioToolkit to extract the HRV signal from the MIT-BIH database. The other steps are described in detail as follows.

Feature extraction: To achieve a more effective combination of features among those shown in literature, 34 features are extracted from the HRV recordings, including 11 linear features, 22 non-linear features and the DC of the heart rate, as shown in Table 1.

Linear parameters: Time domain analysis: The time domain analysis is a statistical method. The HRV signal for each recording is considered the realization of the stochastic process. The HRV signal is measured in the time domain by mathematical statistics. In this work, 7 features of time domain, F1 to F7, are calculated [21].

F1-MEAN: The average of all RR intervals.

F2-RMSSD: The square root value of the mean square of the difference of the adjacent RR intervals.

F3-SDNN: The standard deviation of the RR intervals.

F4-pNN50: The number percentage of heartbeats with a difference of more than 50 ms between the adjacent RR intervals in the whole recording.

F5-CV: The Coefficient of Variation, which is the overall standard deviation (SDNN) divided by the total RR interval average recorded at the same time.

F6-TRI: The HRV triangular index, which is the total number of RR intervals divided by the height of the RR interval histogram.

F7-TINN: The RR intervals' square width of the triangle formed by the histogram.

Frequency domain analysis: Previous studies found that the normal heart rate power spectrum lay between 0 and 0.4 Hz, with a low-frequency band of 0.04-0.15 Hz (*F8-LF*), and a high-frequency band of 0.15-0.4 Hz (*F9-HF*). The total power spectrum between 0-0.4 Hz (*F10-TP*) reflects the HRV size. In addition, the ratio of LF to HF (*F11-LF/HF*) represents the ability of the human autonomic nervous system to regulate the balance.

Non-linear parameters: Many parameters have been used to measure nonlinear properties of HRV, including 1/f scaling of Fourier spectra, the H scaling exponent, and Coarse Graining Spectral Analysis (CGSA). For data representation, there are Poincaré sections, low-dimensional attractor plots, singular value decomposition, and attractor trajectories. The D_2 correlation dimension, Lyapunov exponents, and Kolmogorov entropy have been used as other quantitative descriptions. In this work, 22 nonlinear features, from F12 to F33, were calculated.

F12-ApEn: The Approximate Entropy measures the probability of generating a new pattern in the signal in terms of the complexity of the time series. In this paper, the model dimension m=2 and the similarity tolerance r=0.2 are calculated for the approximate entropy.

F13-SampEn: The Sample Entropy is an improved algorithm with the approximate entropy, which is relatively consistent and more accurate than the approximate entropy and is thus suitable for analysing bio-medical signal sequences.



| No. | Features | | | | | | | |
|-----|---|--|--------------------|--|--|--|--|--|
| F1 | | | MEAN | | | | | |
| F2 | | - | RMSSD√ | | | | | |
| F3 | = | - | SDNN√ | | | | | |
| F4 | _ | Time Domain | pNN50 | | | | | |
| F5 | _ | - | CV√ | | | | | |
| F6 | Linear Parameters | - | TRI | | | | | |
| F7 | _ | - | TINN√ | | | | | |
| F8 | | | LF | | | | | |
| F9 | _ | | HF√ | | | | | |
| F10 | _ | Frequency Domain | TP | | | | | |
| F11 | _ | - | LF/HF [^] | | | | | |
| F12 | | | ApEn | | | | | |
| F13 | _ | - | SampEn√ | | | | | |
| F14 | - | - | DFA ^Δ | | | | | |
| F15 | | | SD ₂ | | | | | |
| F16 | | Poincare Scatter Plot | SD_1/SD_2^{\vee} | | | | | |
| F17 | | | S | | | | | |
| F18 | | | VLI△ | | | | | |
| F19 | | | VAI√ | | | | | |
| F20 | | | CCM△ | | | | | |
| F21 | | | GI | | | | | |
| F22 | Non-Linear | Heart Rate | PI | | | | | |
| F23 | Parameters | Asymmetry | EI | | | | | |
| F24 | | | Dt | | | | | |
| F25 | | | Pa | | | | | |
| F26 | | - | Pb [△] | | | | | |
| F27 | | | Pc | | | | | |
| F28 | | | Pd^ | | | | | |
| F29 | _ | Differential Poincare scatter plot | ED√ | | | | | |
| F30 | _ | | ED1 | | | | | |
| F31 | _ | | ED2 | | | | | |
| F32 | | | ED3 | | | | | |
| F33 | | | ED4 | | | | | |
| F34 | Deceleration Capacity of Heart Rate | Deceleration Capacity of Heart Rate | DC^{\vee} | | | | | |

Table 1: The 34 features extracted in this work.

F14-DFA: A Detrended Fluctuation Analysis is a non-linear method implemented by searching and removing the trend components in the integrated signal at different scales to achieve non-stationary artefacts in the cancellation signal and then extracting the non-stationary signal.

A Poincaré scatter plot [22] is a cross-sectional view of a multidimensional 'spatial structure' with a non-linear blend of properties that is used to observe and study the evolution of nonlinear systems. The Poincaré scatter plot has been applied to the analysis of heart rate variability, and the mapping method is as follows.

For long and short axis measurements, the scatter plot is fitted in an ellipse (the long axis of the ellipse is along the 45-degree contour line), SD_1 is its half-short axis length, $F15-SD_2$ is the half-long axis length, and $F16-SD_1/SD_2$ is the ratio between the half-short axis SD_1 and the half-long axis SD_2 .

F17-S: The elliptical area of the scatter plot.

F18-Vector length index (VLI): The length of the Poincare scatter plot along the 45-degree contour.

F19- Vector-angle index (VAI): The VAI is used to measure the spread of the scatter plot on both sides of the 45-degree contour.

F20-Complex Correlation Measure (CCM): The CCM is calculated using a sliding window in which the time information of the sequence is embedded, which consists of three consecutive points in the scatter plot [23].

Heart Rate Asymmetry (HRA) analysis: In the healthy state, the physiological system is in a non-equilibrium state, and asymmetry is the basic nature of the non-equilibrium system [24]. This asymmetry also exists in a healthy heart system and is known as Heart Rate Asymmetry (HRA). The main measures of heart rate asymmetry (HRA) are the Guzik index (*F21-GI*), Porta index (*F22-PI*) and Ehler index (*F23-EI*).

F21-GI: The distance D to the contours.

F22-PI: The ratio of the number of scatter points below the contour to the number of scatter points that are not on the contour.

F23-EI: The first-order difference transformation of the RR interval sequence to measure the asymmetry of the scatter distribution.

The high dimension time irreversible (HDTI) analysis is used on a Poincaré scatter plot to measure the different delays, the synthesis of the delay in the scatter plot of PI and GI and the new asymmetry index F24- $D_{,c}$.

Another commonly used form of the Poincare Scatter Plot is known as the differential Poincaré scatter plot. The difference scatter plot removes the high correlation of the continuous heartbeat interval compared to the standard scatter plot, and it highlights the interrelationships in continuous cardiac tachycardia (or heart rate) variability. According to the positive and negative values of the difference, the points in the scatter plot will be distributed in the four quadrants of the Cartesian coordinate system.

 $F25-P_{a'}$ $F26-P_{b'}$ $F27-P_{c'}$ and $F28-P_{a'}$ respectively, indicate the ratio between the number of scattered points in A(+,+), B(-,+), C(-,-), and D(+,-) the total number of points.

F29-Distribution Entropy (ED): The distribution entropy reflects the distribution of the scatter points in the differential Poincare scatter plot at different distances from the origin.

 $F30-ED_{1}$, $F31-ED_{2}$, $F32-ED_{3}$ and $F33-ED_{4}$ are respectively, the distribution entropies of the four quadrants in the Poincare scatter plot (including the coordinates of the axis), known as the quadrant distribution entropy.

The coordinates of any point in the scatter plot are composed of two RR interval pairs of the HRV signal time series. The previo s RR interval is the abscissa, and the following RR interval is the ordinate. The interval between the RR intervals is the delay of the scatter plot. A plot with a delay=1 can graphically show the correlation of the adjacent RR interval in the HRV sequential signal.

J Comput Sci Syst Biol, an open access journal ISSN: 0974-7231

Deceleration capacity (DC) of heart rate: F34-DC is determined by following calculation procedures in the literature [15].

DC(AC) = [X(0)+X(1)-X(-1)-X(-2)]/4

The phase-rectified signal averaging (PRSA) signal X (i) is obtained by averaging the signals within the aligned segments; i.e., X(0) is the average of the RR intervals at all anchors, X(1) and X(-1) are the averages of the RR intervals immediately following and preceding the anchors, and so on. This method of quantification corresponds to the quantification of X by a Haar wavelet analysis, for which a scale of 2 is used.

Normalization: In this paper, the Range Scaling method was used to standardize the difference of the eigenvalues of each feature.

$$S_i = \frac{S_i - average}{S_{iMax} - S_{iMin}}$$

This can accelerate the convergence speed of SVM and BPN and improve the classification accuracy while preserving the original distribution characteristics of the data.

Feature selection: Feature selection serves to improve the classification efficiency and accuracy by excluding the noise, redundancy and irrelevant features of the problem by selecting a feature subset of the important information from the original feature set. In this study, two standard feature selection methods are used, the Filter and Wrapper methods. Further, to evaluate the best feature combination, the author uses histogram analysis to select features manually based on the 32 RR-Intervals ECG records.

T-test analysis: The filter method is a type of feature selection

method that is relatively simple. All the characteristics of the evaluation and ranking are used for selection according to the ranking characteristics. In this work, a T-test analysis followed the formula:

$$t = \frac{X_1 - X_2}{\sqrt{\frac{\sigma_{X_1}^2 + \sigma_{X_2}^2 - 2\gamma\sigma_{X_1}\sigma_{X_2}}{n - 1}}}$$

 $\overline{X_2}$, and $\overline{X_2}$ are the average values of two samples. $\sigma_{X_1}^2$ and $\sigma_{X_2}^2$ are the standard deviations of two samples γ is the correlation coefficient.

SVM-RFE: The wrapper feature selection method based on a classifier model is a feature with which the classifier performance can perform evaluations and guide the search process in the feature subset. During the search, this method must call on the training sample classification model to evaluate good or bad characteristics. The process generally requires multiple calls to provide a more accurate evaluation, so the selected feature subsets may become more adapted to subsequent processes. In this work, another feature selection method is Support Vector Machine-Recursive Feature Elimination (SVM-RFE) [25].

Histogram-manual feature selection: A third feature selection method is manual feature selection by a histogram analysis, as shown in Figure 2. The feature distributions of the two groups are illustrated in each histogram. We chose the top 10 features that, according to our experience, show a visible delineation of two groups from the histogram: *F2-RMSSD, F3-SDNN, F5-CV, F6-TINN, F9-HF, F13-SampEn, F16-SD*₁/SD₂, *F19-VAI, F29-ED*, and *F34-DC*. Then, six more features were added: *F11-LF/HF, F14-DFA, F18-VLI, F20-CCM, F26-Pb*, and *F28-Pd*, which were relatively different between the groups.



Automated feature selection: There are large differences in the patterns of the features extracted from ECG records. There may have been an overlap between the classes in the feature space, as Figure 2 shows. Under such a condition, a feature transformation that can minimize the within-class scatter and maximize the between-class scatter will be very useful. In this work, two traditional methods are utilized to reduce the dimension of features: PCA and LDA.

Principal component analysis (PCA): PCA is an unsupervised multivariate analysis method based on the principle of optimal statistics. The original feature space is projected into a new feature space through a series of linear algebraic operations. These new features are irrelevant to each other under the premise of reflecting the original characteristics of the difference information. This process is measured by the covariance matrix of the original feature set. A larger variance indicates a more obvious difference. The dimensionality reduction process of principal component analysis serves to reduce the secondary components that contain less information from the original feature set.

Linear discriminant analysis (LDA): The LDA assumes that intraclass changes exist in a linear subspace of the original feature space, so that the classes are convex and linearly separable. The linear projection method can be used to achieve the linear discriminant of dimensionality. The goal of a linear discriminant analysis is to measure the ratio of the dispersion between-classes and the intra-class dispersion to find the projection direction with the largest ratio. That is, the dispersion of the between-class is made as large as possible, and the dispersion of the intra-class is made as small as possible, which can improve the pattern classification performance.

Arrhythmia classifiers: Many different types of machine learning algorithms have been employed to detect arrhythmia automatically. Of them, SVM and Artificial Neural Network are shown to be good effective classification schemes for arrhythmia detection. To find an easy and fast detection algorithm, in this work, SVM and BPN are exploited.

Support vector machine (SVM): The SVM maps the training sample points of the low-dimensional space into a corresponding high-dimensional space through an appropriate kernel function, so the original linearly inseparable sample points are linearly separable in the high-dimensional space. Then, a hyper-plane is constructed in this high-dimensional space, so as many of the two types of data points as possible are correctly separated. The hyper-plane that shows the largest distance from the nearest data point to the classification surface is the optimal hyper-plane, which is the plane that SVM seeks, and it can ultimately be transformed into a convex quadratic programming problem.

According to the previous experience in the design of an ECG signal classification system, this work chose the RBF kernel function:

 $f(u,v) = e^{-\gamma |u-v|^2}$, γ are the Kernel parameters.

A grid factor is used to determine the penalty factor C and the kernel parameters y.

| | Predicted Condition Positive | Predicted Condition Negative |
|-------------------------|---------------------------------|---------------------------------|
| True Condition Positive | True Positive (TP) | False Negative (FN) |
| True Condition Negative | False Positive (FP) | True Negative (TN) |

Table 2: The confusion matrix for the performance evaluation.

Back propagation neural network (BPN): The BPN is a supervised learning method for training. First, the network is given a set of initial weights, then a sample is entered and the output is calculated. According to the difference between the actual output and the expected output, the appropriate method is used to change the network weight, thereby reducing the difference by repeating the calculation until the difference is less than a predetermined value. A BP neural network consists of an input layer, a hidden layer and an output layer, of which the hidden layer can be extended to multi-layers. This study uses a single hidden layer network.

Performance evaluation: To assess the performance of the classifiers, each classification algorithm was calculated 10 times in an 8-fold cross validation. Then, a confusion matrix (Table 2) containing the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) rates was used to calculate the following performance parameters:

Sensitivity: Sensitivity refers to the ability of a classifier to correctly identify a subject with a normal heart rate.

TPR=TP/(TP+FN)

Specificity: Specificity refers to the ability of the classifier to correctly identify subjects with an abnormal heart rate.

TNR=TN/(TN+FP)

Positive Predictive Value (PPV): When the classifier identifies the subject's heart rate as abnormal, the subject's actual heart rate is abnormal.

PPV=TP/(TP+FP)

Accuracy: The proportion of the correct predictions of the classifier to the total number of tests.

Accuracy=(TP+TN)/(TP+TN+FN+FP)

Computational time: The measured performance using stopwatch timer on a normal PC with CPU 2.4 GHz, RAM 2.0 GB.

Results

The classification performance based on 30 minute ECG recordings: In this study, 96 30-minute ECG recordings are used (normal: arrhythmia=1:1), and 80% was training set and the rest was testing set. 10 times 8-fold cross validation results are shown Tables 3 and 4.

Table 3 shows the arrhythmia classification performance of the SVM classification of different input feature combinations obtained by different methods based on a 30-minute ECG recording. The accuracy of classifications of the training set is similar, and all values are greater than 98.8%. The accuracy range for the test set is from 95.6 \pm 6.5% (by SVM-RFE with LDA) to 98.1 \pm 4.2% (by 34 Original features). In this case, in terms of the computation time, the PCA method is more effective than the T-test with PCA, even though the accuracy is almost the same (97.3 \pm 4.4% by T-test with PCA and 97.1 \pm 4.8% by PCA alone).

Table 4 shows that the classification performance of BPN as a classifier is similar for all 9 input combinations for the training set. However, the classification accuracy for the test set ranges from 92.4 \pm 7.5% (by SVM-RFE with LDA) to 98.3 \pm 3.4% (by the T-test). The input combination of the original 34 features achieves accuracy of 97.6 \pm 4.4%.

The classification performance based on 5 minute ECG recordings

groups. Among the total 576 samples, 80% was used for training and the remaining 20% was used for testing, and carried out 8-fold cross validation 10 times.

During data pre-processing, the 30 minute recordings are segmented into 5 minute recordings, producing 288 (48 recording × 6 pieces) 5 min recording samples for both the normal and arrhythmia

Based on the analysis of the 30-minute recordings, the authors choose the 4 best combinations of algorithms (SVM with 34 original

| | Methods ^{a)} | | Trai | ning | | Testing | | | | | |
|-------|---|-------------------|-------------------|-------------------|------------|-------------------|-------------------|-------------------|------------|--|--|
| No. | | TPR ^{b)} | TNP ^{b)} | PPV ^{b)} | Accuracy | TPR ^{b)} | TNP ^{b)} | PPV ^{b)} | Accuracy | | |
| | | (%) | (%) | (%) | (%) | (%) | (%) | (%) | (%) | | |
| 1 | Original | 100 ± 0 | 99.3 ± 2.2 | 97.8 ± 7.1 | 99.5 ± 1.8 | 100 ± 0 | 96.3 ± 8.4 | 96.9 ± 6.7 | 98.1 ± 4.2 | | |
| 2 | T-test | 100 ± 0 | 99.8 ± 1.2 | 99.4 ± 3.9 | 99.8 ± 1.0 | 97.7 ± 5.8 | 96 ± 8.5 | 96.7 ± 6.9 | 96.9 ± 5.0 | | |
| 3 | SVM-RFE | 100 ± 0 | 99.8 ± 1.2 | 99.4 ± 3.9 | 99.8 ± 1.0 | 98.8 ± 4.4 | 95.4 ± 9.6 | 96.3 ± 7.5 | 97.1 ± 5.0 | | |
| 4 | PCA | 99.6 ± 3.7 | 98.6 ± 3.5 | 95.5 ± 10.6 | 98.8 ± 3.0 | 98.5 ± 4.7 | 95.6 ± 9.1 | 96.4 ± 7.2 | 97.1 ± 4.8 | | |
| 5 | LDA | 99.6 ± 3.7 | 99.9 ± 0.9 | 99.7 ± 2.8 | 99.8 ± 1.0 | 96.7 ± 6.7 | 95.4 ± 10.3 | 96.2 ± 8.2 | 96 ± 6.4 | | |
| 6 | T-test+PCA | 99.2 ± 5.2 | 98.9 ± 2.7 | 96.4 ± 9.3 | 99 ± 2.7 | 97.9 ± 6.1 | 96.7 ± 7.2 | 97.2 ± 6.0 | 97.3 ± 4.4 | | |
| 7 | T-test+LDA | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 97.9 ± 5.5 | 95.2 ± 11.0 | 96.2 ± 8.2 | 96.6 ± 5.9 | | |
| 8 | SVM-RFE+PCA | 98.8 ± 6.4 | 99.2 ± 2.6 | 97.6 ± 7.9 | 99.1 ± 2.4 | 98.1 ± 5.3 | 94.4 ± 11.2 | 95.6 ± 8.6 | 96.3 ± 5.8 | | |
| 9 | SVM-RFE+LDA | 98.3 ± 7.3 | 99.5 ± 2.2 | 98.5 ± 6.9 | 99.3 ± 2.4 | 96.9 ± 6.5 | 94.4 ± 10.9 | 95.4 ± 8.7 | 95.6 ± 6.5 | | |
| a) Fe | a) Feature selection+Dimensional Reduction Method+SVM | | | | | | | | | | |

b) TPR: Sensitivity, TNP: Specificity, PPV: Positive Predictive Value.

Table 3: The classification performance of SVM based on 30 minute recordings.

| | | | Tr | aining | | Testing | | | | |
|--|-----------------------|-------------------|-------------------|-------------------|------------|-------------------|-------------------|-------------------|------------|--|
| No. | Methods ^{a)} | TPR ^{b)} | TNP ^{b)} | PPV ^{b)} | Accuracy | TPR ^{b)} | TNP ^{b)} | PPV ^{b)} | Accuracy | |
| | | (%) | (%) | (%) | (%) | (%) | (%) | (%) | (%) | |
| 1 | Original | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 95.2 ± 8.9 | 96 ± 7.2 | 97.6 ± 4.4 | |
| 2 | T-test | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 96.7 ± 6.7 | 97.1 ± 5.8 | 98.3 ± 3.4 | |
| 3 | SVM-RFE | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 99.4 ± 3.2 | 95 ± 8.6 | 95.8 ± 7.0 | 97.2 ± 4.4 | |
| 4 | PCA | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 99 ± 4.1 | 95 ± 10.4 | 96 ± 8.1 | 97.0 ± 5.5 | |
| 5 | LDA | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 95.4 ± 8.4 | 88.8 ± 14.7 | 90.9 ± 10.9 | 92.1 ± 8.2 | |
| 6 | T-test+PCA | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 98.5 ± 4.7 | 92.9 ± 10.2 | 94.1 ± 8.1 | 95.7 ± 5.5 | |
| 7 | T-test+LDA | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 98.5 ± 4.7 | 93.3 ± 10.5 | 94.5 ± 8.5 | 95.9 ± 5.8 | |
| 8 | SVM-RFE+PCA | 99.8 ± 1.4 | 100 ± 0 | 100 ± 0 | 99.9 ± 0.7 | 98.5 ± 4.7 | 92.7 ± 10.9 | 94 ± 8.6 | 95.6 ± 5.9 | |
| 9 | SVM-RFE+LDA | 100 ± 0 | 99.9 ± 1.1 | 99.8 ± 1.6 | 99.9 ± 0.7 | 96.7 ± 6.7 | 88.1 ± 14.6 | 90.6 ± 10.9 | 92.4 ± 7.5 | |
| a) Feature selection method+Dimensional Reduction Method+BPN | | | | | | | | | | |

b) TPR: Sensitivity, TNP: Specificity, PPV: Positive Predictive Value.

Table 4: The classification performance of BPN based on 30-minute recordings.



Citation: Gong X, Long B, Wang Z, Zhang H, Nandi AK (2018) Faster Detection of Abnormal Electrocardiogram (ECG) Signals Using Fewer Features of Heart Rate Variability (HRV). J Comput Sci Syst Biol 12: 19-27. doi:10.4172/jcsb.1000295

features, SVM with PCA, BPN with 34 original features and BPN with T-test) to analyse the 5-minute recordings. The comparison results are shown in Figure 3. After increasing the number of data sets from 48 to 288 for each group, the best classification accuracy is improved from 98.3% (for the 30 min recording T-test for feature selection classified by BPN) to 99.6% (for the 5 min recording with the original 34 features). Each of the four combinations of algorithms produce a classification accuracy greater than 99.0%.

The classification performance of 32 RR intervals ECG recordings

To achieve a good solution for rapid arrhythmia detection, the third part of this work is a systematic analysis of the classification performance based on 2000 samples of the 32 RR-intervals of the HRV signal. In this study, total 4000 ECG recordings were used (normal: arrhythmia=1:1), and 80% was training set and the rest was testing set, and took 10 times 8-fold cross validation.

Because this is a systematic analysis of arrhythmia detection. Two feature combinations are added. Manual (10) features/ Manual (16) features are selected by a histogram analysis guided by experiences. There are total 11 input feature combinations for both SVM and BPN. The results are illustrated in Table 5. The SVM classifier requires more than 10 times the calculation time for the training set as the BPN classifier. The testing computation time for BPN is approximately twice that of SVM. The fastest computational time is 7.0 ms by Manual (16) features classified by SVM, and it has a 97.2 \pm 0.8% classification accuracy. The best classification accuracy is 97.9 \pm 0.7% by 34 multiple features using BPN, which consumed 22.0 ms.

Based on the 32 RR-Intervals ECG recording, many feature combinations failed to produce classification results. When using BPN as the classifier, 3 input feature combinations fail, including SVM-RFE/SVM-RFE with PCA/SVM-RFE with LDA. When using SVM as the classifier, 6 feature combinations fail, including the T-test/SVM-RFE/T-test with PCA/T-test with LDA/SVM-RFE with PCA/SVM-RFE with LDA. Meanwhile the manual features combinations achieve the similar classification performance as original 34 features combination. For SVM, the manual features combinations even save the computation cost during both training and testing periods.

The comparison between this work with others in the literature

Several researchers are interested in arrhythmia detection by ECG signals directly or by analysing the HRV signals. In Table 6, the comparison results are summarized between the two groups Arrhythmia and Normal groups proposed in the literature and this work. To find the better features for portable ECG equipment, a two-group classification is considered in this study. In the literatures Elif's [7], Jenerfar's [9], and Nuryani's [10] works both focus on QRS wave analysis based on different time size of HRV signals. In this work, the best classification performance results of two groups is 99.6% when using the 34 original features based on the 5 minute ECG records. Compared with manually selected 16-feature combination, the manually selected 10-feature combination performs a little better whenever used with SVM or BPN as the classifier. The manual-selected 10 features combination holds the similar classification performance (97.4%) as the 34 original features (97.9%) with a fast testing time (about 10.0 ms) based on 32 RR-Intervals ECG records.

| | | SVM | | | | BPN | | | | |
|-----|----------------------|------------|---------------|------------|------------|------------|---------------|----------|------------|-------------|
| No | Method | Training | | Testing | | Training | | | Testing | |
| NO. | | Accuracy | Time | Accuracy | Time | Accuracy | Time | Accuracy | Tir | ne |
| | | (%) | (S) | (%) | (ms) | (%) | (S) | (%) | (m | is) |
| 1 | Original | 99.4 ± 0.1 | 1704.8 ± 12.0 | 97.9 ± 0.8 | 12.0 ± 2.0 | 98.8 ± 2.6 | 4.0 ± 1.2 | | 97.9 ± 0.6 | 22.0 ± 2.0 |
| 2 | T-test | - | - | - | - | 100 ± 0 | 3.3 ± 0.9 | | 97.8 ± 0.7 | 22.0 ± 1.0 |
| 3 | SVM-RFE | - | - | - | - | - | - | | - | - |
| 4 | PCA | 96.9 ± 0.9 | 946.6 ± 3.2 | 94.4 ± 1.5 | 13.0 ± 4.0 | 96.9 ± 4.4 | 75.7 ± 0.8 | | 94.0 ± 1.0 | 21.0 ± 1.0 |
| 5 | LDA | 96.4 ± 0.8 | 1264.7 ± 10.2 | 95.7 ± 1.4 | 8.0 ± 2.0 | 96.9 ± 4.4 | 78.0 ± 1.0 | | 95.2 ± 1.4 | 21.0 ± 0.0 |
| 6 | T-test+PCA | - | - | - | - | 93.8 ± 5.1 | 77.4 ± 0.6 | | 93.8 ± 1.0 | 21.0 ± 1.0 |
| 7 | T-test+LDA | - | - | - | - | 95.0 ± 4.9 | 77.4 ± 0.9 | | 95.5 ± 1.5 | 21.0 ± 1.0 |
| 8 | SVM-RFE+PCA | - | - | - | - | - | - | | - | - |
| 9 | SVM-RFE+LDA | - | - | - | - | - | - | | - | - |
| 10 | Manual (10) features | 99.0 ± 0.4 | 596.0 ± 2.8 | 97.4 ± 0.8 | 10.0 ± 4.0 | 98.8 ± 2.6 | 146.6 ± 4.3 | | 97.1 ± 0.5 | 20.0 ± 1.0 |
| 11 | Manual (16) features | 98.2 ± 0.7 | 952.8 ± 5.5 | 97.2 ± 0.8 | 7.0 ± 1.0 | 98.8 ± 2.6 | 66.8 ± 6.2 | | 97.0 ± 1.0 | 28.0 ± 15.0 |

 Table 5: The training and testing performances and the time for testing for different classification algorithms based on the 32 RR-Intervals recording data.

| No. | Author | Methods | Classes | Accuracy (%) | Testing time |
|-----|---------------------|---|---------|--------------|--------------|
| | | (Data set+Feature selection+Classifier) | | | (ms) |
| 1 | This work | 5-min HRV+34 Features+BPN | 2 | 99.6 | - |
| 2 | Elif et al. [7] | 1 beat record+QRS-DWT+SVM | 2 | 99.4 | - |
| 3 | This work | 30-min HRV+T-test+BPN | 2 | 98.3 | - |
| 4 | This work | 32RR HRV+34 Features+BPN | 2 | 97.9 | 22.0 ± 2.0 |
| 5 | Nuryani et al. [10] | Difference between-two RR intervals+SVM | 2 | 97.5 | - |
| 6 | This work | 32 RR HRV+10 Features+SVM | 2 | 97.4 | 10.0 ± 4.0 |
| 7 | This work | 32 RR HRV+16 Features+SVM | 2 | 97.2 | 7.0 ± 1.0 |
| 8 | Jenefar et al. [9] | 10-min HRV-QRS+SVM | 2 | 88 | - |

Table 6: A comparison of the present study and the literature for the 2 classes of Abnormal and Normal.

Citation: Gong X, Long B, Wang Z, Zhang H, Nandi AK (2018) Faster Detection of Abnormal Electrocardiogram (ECG) Signals Using Fewer Features of Heart Rate Variability (HRV). J Comput Sci Syst Biol 12: 19-27. doi:10.4172/jcsb.1000295

Discussion

To adapt to the online screening of arrhythmia, a simple and rapid algorithm is required. The aim of this work is to find better combinations of features for distinguishing abnormal and normal ECG records [26]. Tsipouras et al. [27] calculated different characteristics of the HRV signals to serve as the classification features. Compared with the morphological classification, a classification based on an analysis of HRV features appears to be effective for noise suppression, even for very short recordings of ECG raw data (i.e., a 32 RR intervals). An investigation of those features is discussed by Babak et al. [13] who classified 6 subclasses of arrhythmia with a performance accuracy of approximately 99.6%. Compared with 15 features used in Babak's work [13]. This study increased the number of features to 34 to include linear features, nonlinear features and the DC of the heart rate, which provide useful information about the ECG signal for detecting arrhythmia. The 34 original features employed in this work are effective and stable as the input for detecting arrhythmia in 3 different data sizes of ECG recordings.

During this systematic comparison, the classification accuracy and computational time are considered. In this study, both the SVM and BPN, the testing computational time is only several milliseconds based on short ECG records. When some features are distorted, compared with SVM, BPN is more stable. Based on the same input feature combination, BPN requires shorter training computational time than SVM does.

To find the best and simplest feature combination, it is investigated the influence of different feature combinations as well as the computational time for both training and testing sets. When an ECG recording is short (32-RR intervals), some feature combinations fail to separate the two groups. The 10 feature combinations manually selected by a histogram (*F2-RMSSD*, *F3-SDNN*, *F5-CV*, *F6-TINN*, *F9-HF*, *F13-SampEn*, *F16-SD*₁/*SD*₂, *F19-VAI*, *F29-ED*, and *F34-DC*) achieves the optimal classification performance, similar to the 34 original features. The additional 6 features (*F11-LF/HF*, *F14-DFA*, *F18-VLI*, *F20-CCM*, *F26-Pb*, and *F28-Pd*) do not improve the classification performance very well. These manually selected features might be the most important features to distinguish an arrhythmia from a normal ECG. They include linear features, nonlinear features and the DC of the heart rate. This 10-manual-selected features combination may be used in the future to faster detect the Arrhythmia with fewer features.

Conclusion

A systematic performance comparison of combinations of 11 input features selected by different methods and 2 classifiers (SVM and BPN) is performed. BPN requires a shorter training computational time than SVM but offered a similar classification performance. Based on raw ECG recordings of different durations, the 34 original features proposed in this paper had a very stable classification performance, and the best performance of 99.6% was achieved by BPN using 5-minute ECG recording. We discovered that the combination of 10 manually selected features (RMSSD, SDNN, CV, TINN, HF, SampEn, SD1/ SD2, VAI, ED, and DC) provides relevant information for detecting arrhythmia and can achieve the same classification performance as can the 34 original features with 32 RR-Intervals ECG recordings. It takes several milliseconds of the testing computational time and several seconds of the training computational time. These manual-selected 10 features used by SVM classifier is expected to be suitable for portable ECG monitoring equipment.

Acknowledgements

Professor Nandi is a Distinguished Visiting Professor at Tongji University, Shanghai, China.

Funding

This work was partly supported by the National Science Foundation of Shanghai grant number 16JC1401300 and the Fundamental Research Funds for the Central Universities of Tongji University (22120170043).

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

References

- World Health Organization (2018) Cardiovascular Diseases (CVDs). Available from: http://www.who.int/ediacentre/factsheets/fs317/en/
- Serkan K, Turker I, Jenni P (2011) Personalized long-term ECG classification: A systematic approach. Expert Syst Appl 38: 3220-3226.
- Paola M, Remi D, Pierre R, Christine H (2011) Arrhythmia discrimination in implantable cardioverter defibrillators using support vector machines applied to a new representation of electrograms. IEEE Biomed Eng 58: 1797-1803.
- Khorrami H, Moavenian M (2010) A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. Expert Syst Appl 37: 5751-5757.
- Ceylan R, Ozbay Y (2007) Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. Expert Syst Appl 33: 286-295.
- Martis RJ, Acharya UR, Prasad H, (2013) Automated detection of atrial fibrillation using Bayesian paradigm. Knowl Based Syst 54: 269-275.
- Elif DU (2008) Support vector machines for detection of electrocardiographic changes in partial epileptic patients. Eng Appl Artif Intel 21: 1196-1203.
- Alonso-Atienza F, Morgado E, Fernández-Martínez L (2014) Detection of lifethreatening arrhythmias using feature selection and support vector machines. IEEE Trans Biomed Eng 61: 832-840.
- Jenefar SC, Vanitha L (2014) Prediction of Sudden Cardiac Death Using Support Vector Machine. International Conference on Circuit. Power and Computing Technologies, Penang, Malaysia, pp: 377-381.
- Nuryani N, Harjito B, Yahya I, Lestari A (2015) Atrial fibrillation detection using support vector machine. Proceedings of the Joint International Conference on Electric Vehicular Technology and Industrial, Mechanical, Electrical and Chemical Engineering (ICEVT & IMECE), Surakarta, Indonesia, pp: 215-218.
- Rouhani M, Soleymani R (2009) Neural Networks based Diagnosis of heart arrhythmias using chaotic and nonlinear features of HRV signals. International Association of Computer Science and Information Technology-Spring Conference, pp: 545-549.
- Shaffer F, Ginsberg JP (2017) An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health 5: 258.
- Babak MA, Seyed KS, Maryam M (2008) Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. Artif Intell Med 44: 51-64.
- Axel B, Kantelhardt JW, Barthel P, Schneider R, Mäkikallio T, et al. (2006) Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. Lancet 367: 1674-1681.
- Bauer A, Kantelhardt JW, Barthel P (2006) Deceleration capacity of heart rate as a predictor of late mortality after myocardial infarction. Heart Rhythm 3: 1674-1681.

J Comput Sci Syst Biol, an open access journal ISSN: 0974-7231

Citation: Gong X, Long B, Wang Z, Zhang H, Nandi AK (2018) Faster Detection of Abnormal Electrocardiogram (ECG) Signals Using Fewer Features of Heart Rate Variability (HRV). J Comput Sci Syst Biol 12: 19-27. doi:10.4172/jcsb.1000295

- Martis RJ, Acharya UR, Adeli H (2014) Current methods in electrocardiogram characterization. Comput Biol Med 48: 133-149.
- Acharya UR, Fujita H, Lih OS, Hagiwara Y (2017) Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. Inf Sci 415: 190-198.
- He Z, Zhang X, Cao Y, Liu Z, Zhang B, et al. (2018) LiteNet: Lightweight Neural Network for Detecting Arrhythmias at Resource-Constrained Mobile Devices. Sensors 18: 1229-1237.
- 19. Moody GB, Mark RG (2001) The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 20: 45-50.
- 20. Physionet (2018) Available from: http:// www.physionet.org
- 21. Ho K, Moody GB, Peng CK, Mietus J, Larson M, et al. (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: 842-848.

- 22. Khandoker AH, Karmakar C, Brennan M, Voss A (2013) Poincaré Plot Interpretation of HRV Using Physiological Model. Poincaré Plot Methods for Heart Rate Variability Analysis. Springer, US, pp: 25-46.
- Karmakar CK, Khandoker AH, Gubbi J, Palaniswami M (2009) Complex Correlation Measure: a novel descriptor for Poincare plot. Biomed Eng 8: 1-12.
- 24. Chialvo DR, Millonas MM (1995) Asymmetric Unbiased fluctuations are sufficient for the operation of a correlation ratchet. Phys Lett A 209: 26-30.
- Zhou X, Tuck DP (2007) MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. Bioinformatics 23: 1106-1114.
- Guo Y, Hastie T, Tibshirani R (2007) Regularized linear discriminant analysis and its application in microarrays. Biostatistics 8: 86-100.
- Tsipouras MG, Fotiadis DI (2004) Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability. Comput Meth Prog Bio 74: 95-108.