Coronary artery disease (CAD) is a leading cause of death worldwide. Heart rate variability (HRV) has been proven to be a non-invasive marker of the autonomic modulation of the heart. Nonlinear analyses of HRV signals have shown that the HRV is reduced significantly in patients with CAD. Therefore, in this work, we extracted nonlinear features from the HRV signals using the following techniques: recurrence plots (RP), Poincare plots, and detrended fluctuation analysis (DFA). We also extracted three types of entropy, namely, Shannon entropy (ShanEn), approximation entropy (ApEn), and sample entropy (SampEn). These features were subjected to principal component analysis (PCA). The significant principal components were evaluated using eight classification techniques, and the performances of these techniques were evaluated to determine which presented the highest accuracy in classifying normal and CAD classes. We observed that the multilayer perceptron (MLP) method resulted in the highest classification accuracy (89.5%) using our proposed technique.

**Keywords**: Coronary artery disease; heart rate variability; feature extraction; feature significance; principal component analysis; supervised classification.
built-up plaque ruptures, limiting blood flow to the heart muscle. CAD is the most common cause of sudden cardiac death. Therefore, early detection of this condition could help reduce the alarming mortality rate associated with this disease.

The electrocardiogram (ECG), a technique that records the electrical activity of the heart over time, is commonly used for detecting CAD. However, 50% to 75% of patients with CAD do not show symptoms in ECGs. Another technique, the exercise stress test, has limited use because of its low sensitivity and because not all patients can reach the target heart rate required for the test. Several other clinical diagnostic techniques, such as single-photon emission computed tomography, radionuclide exercise myocardial perfusion imaging, positron emission tomography, cardiac magnetic resonance imaging, and multi-slice computed tomography are also available. Each of these techniques has at least one limitation, such as lower sensitivity, specificity, higher cost, longer examination time, and/or injection of radioactive material. Thus, the development and research of non-invasive diagnostic techniques are becoming more extensive because these techniques are less expensive, safer, and more accessible to a wide range of patients. Non-invasive techniques are also more capable of detecting CAD at early stages when luminal obstructions are not visible yet.

One such non-invasive technique makes use of is the heart rate variability (HRV), which measures the variation in the heart rate from beat to beat. HRV has proven to be a non-invasive marker of the autonomic modulation of heart. Using this method, the heart rate (HR) time series is computed by measuring time intervals between consecutive R peaks \( t_{r-r} \), expressed in seconds, of the QRS complex of the ECG signal. The resulting time series is referred to as an RR interval time series. The sequence of numbers that results is the series of instantaneous time periods, and the instantaneous HR, expressed in beats per minute, is computed as \( HR = 60/t_{r-r} \). A plot of the instantaneous HR gives the HRV signal. The assessment of HRV has been shown to aid in the clinical diagnosis of and intervention strategies for treating cardiological and non-cardiological diseases, including dilated cardiomyopathy, congestive heart failure, fetal distress conditions, obstructive sleep apnea, stroke, depression, other psychological factors, and diabetic neuropathy.

A decrease in the HRV indicates some type of cardiac or autonomic dysfunction. Reduced baroreflex sensitivity has been shown to be related to CAD, and reduced vagal activity, i.e., altered HRV, has been observed in patients with CAD. The circadian rhythm of HRV has been demonstrated to be reduced in patients with CAD. The reduction of low-frequency power has also been observed to correlate with the angiographic severity of CAD. Spectral analysis indicating a reduction in high-frequency power has also been reported. The time and frequency domain measures of HRV are lower in patients with chronic or subacute CAD than in healthy subjects. However, these methods are sensitive to noise and hence, may not always provide an accurate and clear categorization of healthy and CAD patients. A noise detection and elimination strategy for acoustic detection in CAD was proposed by Zia et al. The cardiovascular system is complex, and studies performed under...
the assumption that it is a nonlinear system have shown that HRV is significantly reduced in patients with CAD. Therefore, many studies have used nonlinear techniques to analyze the HRV signal. For comparison, a more detailed presentation of the results obtained in other CAD studies is given in the discussion section of this paper.

For this work, we extracted nonlinear features from the HRV signals using the following techniques: recurrence plots (RP), Poincare plots, and detrended fluctuation analysis (DFA). We also extracted three types of entropy, namely, Shannon entropy (ShanEn), approximation entropy (ApEn), and sample entropy (SampEn). These features were subjected to principal component analysis (PCA) to extract significant orthogonal features derived from the existing feature set. These significant principal components were later subjected to eight classification techniques, and the performances of these techniques were evaluated to determine which classifier presents the highest accuracy for classifying samples to the corresponding normal and CAD classes. The paper is organized as follows: Section 2 contains a description of the data used for this work. Section 3 contains a brief description of our preprocessing method. The methodology is given in Sec. 4, and the subsequent explanation of the results and a comparison with the previous literature can be found in Sec. 5. Our discussion is located in Sec. 6, and our conclusions are provided in Sec. 7.

2. Data

The ECG signals were recorded from patients with CAD and healthy volunteers, using BIOPAC™ equipment and converted to heart rate time series using ACQ Knowledge software. The sampling rate was 500 Hz. Both groups ranged from 40 to 70 years in age, and the mean age was 55 years. The CAD group comprised 10 Iqraa Hospital (Calicut, Kerala, India) in-patients who were under treatment for CAD. The cardiologist identified these patients and as their medications were similar, the drug effects on the HRV signal were also assumed to be similar. Figures 1(a) and 1(b) show the sample signals obtained from a CAD patient and a healthy volunteer, respectively. In this work, we collected data from 10 normal and 10 CAD subjects. From each subject, we recorded 1,000 samples. Thus, we obtained 61 datasets of normal and 82 datasets of CAD, each containing 1,000 samples from 10 subjects.

3. Pre-Processing

The original ECG signals were passed through a second order Butterworth high-pass filter with a cut-off frequency of 0.3 Hz to suppress baseline wander and a low-pass filter (second order Butterworth filter) of cut-off frequency of 15 Hz to remove unwanted high frequencies present in the ECG signal. Finally, the signal was fed to a median filter to extract the baseline wander of the processed ECG signal, and then subtracted from the processed ECG signal to effectively remove all the baseline wander. The R peaks of ECG signal were detected using Tompkins’s algorithm.
4. Methodology

This section contains brief descriptions of the extracted nonlinear features and PCA. It also provides a description of the eight classifiers that were evaluated.

4.1. Feature extraction

The following section contains a description of the different feature extraction techniques employed.

4.1.1. Recurrence plots (RP)

Recurrence quantification analysis (RQA) is a popular method used for nonlinear data analysis. In a state space trajectory of a dynamical system, RQA can be used to quantify the number and duration of recurrences. A cross recurrence plot (CRP) reveals all the times at which the phase space trajectory of the first dynamical system visits approximately the same area in the phase space that the phase space trajectory of the second dynamical system is located. Webber and Zbilut, Zbilut and Webber, and Marwan et al. developed measures for the quantification of RP. Some of these measures include recurrence rate (REC), determinism (DET), laminarity (LAM), mean diagonal line length ($L_{\text{mean}}$), trapping time ($TT$), longest diagonal line ($L_{\text{max}}$), longest vertical line ($V_{\text{max}}$), and entropy (ENTR). In this work, the RQA features found to be significant are REC, DET, and $L_{\text{mean}}$. These features are defined as follows. In these definitions, $N$ represents the number of points $R_{i,j}$ on the phase space trajectory, $N_r$ represents the number of vertical lines in the recurrence plot, $N_l$ represents the number of diagonal lines in the RP, and $P(l)$ and $P(v)$ represent the histogram of the line lengths of the diagonal and vertical lines, respectively. Figure 2 shows the typical RP of CAD and normal heart plots.

Fig. 1. HRV signals of (a) a CAD patient and (b) a healthy volunteer.
The recurrence rate is the percentage or density of the recurrence points in a recurrence plot and is given by

$$REC = \frac{1}{N^2} \sum_{i,j=0}^{N} R_{i,j}. \quad (1)$$

Determinism is the fraction of recurrence points forming diagonal lines of minimal length $l_{\text{min}}$. DET indicates the predictability of the system, as denoted below:

$$DET = \frac{\sum_{l=l_{\text{min}}}^{N} lP(l)}{\sum_{i,j} R_{i,j}}. \quad (2)$$

The mean diagonal line length is the mean length of the diagonal lines and is given by

$$L_{\text{mean}} = \frac{\sum_{l=l_{\text{min}}}^{N} lP(l)}{\sum_{l=l_{\text{min}}}^{N} P(l)}. \quad (3)$$

4.1.2. Poincare plots

The Poincare plot, a technique taken from nonlinear dynamics, portrays the nature of RR interval fluctuations. In this technique, each RR interval is plotted as a function of the previous RR interval. This plot may be analyzed quantitatively by calculating the standard deviations of the distances of the $R-R(i)$ to the lines $y = x$ and $y = -x + 2R - R_{m}$, where $R-R_{m}$ is the mean of all $R-R(i)$. The standard
deviations are referred to as SD1 and SD2, respectively. SD1 is related to the fast beat-to-beat variability in the data, while SD2 describes the longer-term variability of $R-R(i)$. Figure 3 shows the typical recurrence plots of CAD and normal heart plots.

4.1.3. Detrended fluctuation analysis (DFA)

Detrended fluctuation analysis (DFA) is used to quantify the fractal scaling properties of short RR interval signals. This technique is a modification of the root-mean-square analysis of random walks applied to non-stationary signals. The root-mean-square fluctuation of an integrated and detrended time series $F(n)$ is measured at different observation windows and plotted against the size of the observation window $n$ on a log–log scale. Two parameters are determined from this technique. The first of these parameters is the slope of the line relating $\log F(n)$ to $\log n$ (called $\alpha$); the second parameter is the dimension of the time series ($D$).

4.1.4. Approximate entropy (ApEn)

Approximate entropy (ApEn), proposed by Acharya et al. and Piscus, is the logarithmic likelihood that the trends of the data patterns that are close to each other will remain close when compared with the next pattern. Thus, ApEn is a measure of data regularity. A greater likelihood of high regularity results in smaller ApEn values, and vice versa. Consider a time series $x(n), n = 1, 2, \ldots, N$. A series of patterns of length $e$ (called the embedding dimension which is the smallest integer for which the patterns do not intersect with each other) is derived from $x(n)$. ApEn is given by

$$ ApEn(e, r, N) = \frac{1}{(N - e + 1)} \sum_{i=1}^{N-e+1} \log C_i^e(r) - \frac{1}{(N - e)} \sum_{(i-1)}^{N-e} \log C_i^{e+1}(r), \quad (4) $$
where the index $r$ is a fixed parameter which sets the *tolerance* of the comparison and $C^e_i(r)$ is the correlation integral given by

$$C^e_i(r) = \frac{1}{(N - e + 1)} \sum_{j=1}^{N-e+1} \Theta(r - ||x_i - x_j||).$$

(5)

For this work, $r$ was set to be 0.2 times the standard deviation of the time series, and $e$ was chosen to be 10.

4.1.5. Sample entropy (SampEn)

Sample entropy (SampEn), proposed by Richman and Moorman, is the negative natural logarithm of an estimate of the conditional probability that patterns of length $e$ that match point-wise within a tolerance $r$ also match at the next point.28 Larger values of SampEn correspond to more irregularity in the data and vice versa. In sample entropy calculation, runs of points matching within the tolerance $r$ are performed until there is no match, while the count of template matches are stored in counters $A(k)$ and $B(k)$ for all lengths $k$ up to $e$. SampEn is given by the formula

$$\text{SampEn}(k, r, N) = \ln \frac{A(k)}{B(k - 1)},$$

(6)

where $k = 0, 1, \ldots, e-1$ and $B(0) = N$, the length of the input series.

4.1.6. Shannon’s entropy (ShanEn)

Spectral entropy (SEN), a normalized form of Shannon’s entropy, uses the amplitude components of the power spectrum of the EEG signal for entropy calculations.29,30 It quantifies the spectral complexity of the time series. Entropy is computed by multiplying the power in each frequency $p_f$ by the logarithm of the same power and then multiplying the result by $-1$. The total ShanEn entropy is the sum of entropy computed over the entire frequency range31 and is given by

$$\text{ShanEn} = \sum_f p_f \log \frac{1}{p_f}.$$  

(7)

4.1.7. Short-range scaling exponent ($\alpha 1$)

The total length of the HR signal ($N$) is integrated

$$y(k) = \sum_{i=1}^{k} [HR(i) - HR_{\text{avg}}],$$

(8)

where $HR(i)$ is the $i$th heart rate signal and $HR_{\text{avg}}$ is the average heart rate of $N$ samples. Next, the integrated time series is divided into boxes of equal length $n$, a least-squares line is fit to the data. The $y$-coordinate of the straight line segments is given by $y_n(k)$. The root mean square fluctuation of this integrated and detrended
heart rate data is calculated by

\[ F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2}. \] (9)

This computation is repeated over all the box sizes to provide a relationship between \( F(n) \), and box size \( n \). In this study, the box size is ranged from 4 to \( \sim 300 \) beats.

A linear relationship on a double-log graph indicates the presence of scaling \((F(n) \approx n^\alpha)\). The slope of the line relating \( \log F(n) \) to \( \log(n) \) to depicts the scaling exponent (\( \alpha \)) which describes the correlation properties of the entire heart rate series. The short-range scaling exponent (\( \alpha_1 \)) was evaluated over periods of 4 to 13 beats. The value of \( \alpha_1 \) implies the fractal behavior of the time series.

4.1.8. Correlation dimension (D2)

\( D_2 \) is defined to obtain a quantitative measure of the nature of trajectory, and the ranges of \( D_2 \) are different for various heart diseases.\(^6\) The \( D_2 \) of the attractor is given by:

\[ D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log(r)}, \] (10)

where the correlation integral \( C(r) \) is given by

\[ C(r) = \frac{1}{N^2} \sum_{x=1}^{N} \sum_{y=1, x \neq y}^{N} \Theta(r - |X_x - X_y|), \] (11)

where \( X_x, X_y \to \) points of the trajectory in the phase space,

\( N \to \) is the number of data points in phase space,

\( r \to \) radial distance around each reference point \( X_i \), and

\( \Theta \to \) is the Heaviside function.

4.2. Principal components analysis

Principal components analysis (PCA) is a feature extraction technique that transforms a high-dimensional dataset to a low-dimensional orthogonal feature (eigen-vector) space while retaining the maximum variance of the original high dimensional dataset. Typically, each orthogonal feature is referred to as a principal component. The first principal component captures the most significant variance in the dataset. The second principal component is directionally perpendicular to the first principal component that contains the next significant variance with respect to the first principal component. Principal components are ranked by their corresponding Eigenvalues, which are scalar representations of the degree of variance within the corresponding principal components.
Given an $n$-dimensional numeric dataset, each dimension represents a coordinate of the $n$-dimensional data space. Our objective here is to transform the coordinate space from the $n$-dimensional space to another desired coordinate data space (of lower dimension), based on the underlying distribution of data points in the dataset.

With the application of PCA, the subsequent set of coordinates for each data point is generated to represent the underlying variance distribution. In Fig. 4, we present the PCA workflow, which includes the following steps:

1. **Calculating covariance matrix**: Subtract the data mean in all dimensions to produce a data set with a zero mean, and then calculate the covariance matrix.

2. **Extracting eigenvalues and eigenvectors**: Subject the covariance matrix to decomposition that results in a matrix of eigenvectors in an $n$-dimensional space, which consists of a set of $n$ principal components in $n$ dimensions, and their corresponding eigenvalues.

3. **Sorting eigenvectors**: Rank the eigenvectors according to eigenvalue, starting from the highest value and moving to the lowest value.

4. **Selecting significant eigenvectors**: Select the most significant eigenvectors based on their corresponding eigenvalues, and project the given data on the selected eigenvectors (or principal components).

In the above steps, the number of selected principal components depends on the percentage of the total variances reflected by their corresponding Eigenvalues. The total variance of the dataset is equal to the sum of all eigenvalues. In this work, the original CAD dataset consisted of nine significant nonlinear features that are listed in the next section (Table 1). We subjected these nonlinear features to PCA and selected the six most significant principal components which account for 95% of the variance of the nonlinear-feature dataset. The obtained principal components are shown in Table 2.

### 4.3. Classifiers

The following section enumerates and provides a brief overview of the classifiers employed in this study.

#### 4.3.1. Bayesian network (BN)

A Bayesian network (BN) is a graphical model that encodes probabilistic relationships among the features. A BN consists of a set of features and a set of directed edges.
between the features that form a directed acyclic graph and a conditional probability table, which represents the uncertainty of relationships between the features and the respective parents. Two steps are involved when a BN is used for classification: learning the BN structure and learning the parameters for the structure.33

### 4.3.2. Naive Bayes (NB)

The Naive Bayes classifier is a probabilistic classifier based on the Bayes theorem that has a strong assumption that the features are independent random variables.34 This assumption helps the classifier to compute the probabilities (probability that a given tuple belongs to a particular class) required by the Bayes formula.

### 4.3.3. Support vector machine (LibSVM)

Support vector machines are another set of related supervised learning methods used for classification. An SVM will construct a separating hyper-plane that maximizes the margin between the input data classes which are viewed in an n-dimensional space. To calculate the margin, two parallel hyper-planes are constructed, one on each side of the separating hyper-plane. These two hyper-planes are computed directly using the training set. We chose SVM because of its superior generalization in high-dimensional data and fast convergence in training.35

---

**Table 1.** Range of variance significant nonlinear features for the normal and CAD groups.

<table>
<thead>
<tr>
<th>Features</th>
<th>Normal (Mean ± SD)</th>
<th>CAD (Mean ± SD)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 SD2</td>
<td>64.8 ± 30.1</td>
<td>41.2 ± 26.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2 Lmean</td>
<td>12.7 ± 6.67</td>
<td>23.0 ± 17.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3 REC</td>
<td>34.9 ± 11.1</td>
<td>49.5 ± 18.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>4 DET</td>
<td>98.0 ± 1.42</td>
<td>99.0 ± 1.13</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>5 ShanEn</td>
<td>3.22 ± 0.388</td>
<td>3.64 ± 0.560</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6 ApEn</td>
<td>1.16 ± 0.188</td>
<td>1.01 ± 0.228</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>7 SampEn</td>
<td>1.53 ± 0.335</td>
<td>1.22 ± 0.476</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>8 α1</td>
<td>1.14 ± 0.235</td>
<td>0.865 ± 0.383</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>9 D2</td>
<td>2.36 ± 1.44</td>
<td>0.349 ± 0.416</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**Table 2.** Range of variance of six principal components obtained using 0-1 normalization and PCA accounting for 95% of variance.

<table>
<thead>
<tr>
<th>Principal components</th>
<th>Normal (Mean ± SD)</th>
<th>CAD (Mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.01889 ± 0.320712</td>
<td>-0.48805 ± 0.349908</td>
</tr>
<tr>
<td>2</td>
<td>0.262432 ± 0.232432</td>
<td>0.011774 ± 0.225043</td>
</tr>
<tr>
<td>3</td>
<td>-1.34019 ± 0.154243</td>
<td>-1.13456 ± 0.266746</td>
</tr>
<tr>
<td>4</td>
<td>0.034841 ± 0.19559</td>
<td>0.031956 ± 0.223662</td>
</tr>
<tr>
<td>5</td>
<td>-0.173088 ± 0.132831</td>
<td>0.255909 ± 0.124357</td>
</tr>
<tr>
<td>6</td>
<td>-0.41429 ± 0.111811</td>
<td>-0.45782 ± 0.102175</td>
</tr>
</tbody>
</table>
4.3.4. **Sequential minimal optimization (SMO)**

Sequential minimal optimization (SMO) is an algorithm that increases the speed of SVM training.\(^\text{36}\) We implemented the Platt’s SMO algorithm for this work.

4.3.5. **Artificial neural network (ANN)—multilayer perceptron (MLP)**

For this work, the ANN—MLP classifier was trained using the back propagation algorithm. ANN—MLP works in the following way: once the network is designed and weights are initialized, the training by this algorithm is performed in three stages — the feed-forward of the input features, the back-propagation of the error between the network output and desired output, and the adjustment of the weights. These stages are repeated iteratively until the mean square error between the network output and desired output is less than a pre-set threshold. The weight update is used to maximize the rate of error reduction, and, hence, it is called a “gradient descent” algorithm.\(^\text{37}\)

4.3.6. **Radial basis function (RBF)**

An RBF network is a type of ANN algorithm that generally has three layers — an input layer, a hidden layer with a non-linear RBF activation function, and a linear output layer. A radial basis function is a real-valued function that has a value that depends only on the distance from the origin or from a center. Such a function \(\phi\) should also satisfy the relationship \(\phi(x) = \phi(||x||)\).

4.3.7. **Random forest**

The random forest classifier uses a group of decision trees and generates class labels for data points by combining the outputs of the individual trees. Each tree is grown using a subset of the possible features. During classification, each tree “votes” for one of the classes, and the most popular class is assigned.\(^\text{38,39}\)

4.3.8. **Classification and regression trees (CART)**

CART is a non-parametric decision tree-based learning technique. The model produces a classification tree if the target variable is categorical or a regression tree if the target variable is numeric. In our work, a classification tree was produced and then used for classification.

5. **Results**

This section contains a presentation of the range of nonlinear features obtained from both classes of data and the subsequent classification results.

5.1. **Significant nonlinear features and extracted PCA features**

The dataset comprised 61 HRV signals from the normal group and 82 from the CAD group. The nonlinear features described in the previous section were extracted from
these signals. The t-test was applied to select those features that were significantly different from the others. We determined that a p-value of less than 0.0001 indicates that the feature is significant. Table 1 lists the variance (Mean ± Standard Deviation (SD)) of the significant nine features for the two groups. It can be seen that these features have a p-value of less than 0.0001. Subsequently, PCA was employed on these features (after 0-1 normalization), and the top six principal components were determined. The top six principal components accounted for 95% of the variance. Table 2 presents the range of variance captured by these principal components for the normal and CAD groups.

5.2. Classification results

The eight classification techniques that were described in the previous section were evaluated on the selected features to compare the performance of these techniques. We employed five-fold cross validation to select training and testing datasets. The classifiers studied were BN, NB, LibSVM, ANN-MLP, RBF, SMO, random forest, and CART. In Table 3, we list the settings of parameters for each classification method.

We evaluated the performance of classifiers using the overall classification accuracy and the true positive rate (TPR) for diagnosing CAD. Accuracy is the proportion of the correctly classified samples from the total number of samples. TPR or sensitivity is the probability that the classifier will predict a positive result for a patient with CAD. In Figs. 5 and 6, we illustrate the classification accuracies and TPR of the CAD recognition obtained using eight classifiers for easy comparison. We consider those classification methods that have both evaluation metrics ranking in

<table>
<thead>
<tr>
<th>Methods</th>
<th>Parameters used</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>Estimator algorithm: Simple Estimator with alpha = 0.5; Algorithm for searching network structure: Hill climbing</td>
</tr>
<tr>
<td>NB</td>
<td>Minimum number of instances: 0; Use Kernel estimator: No; Use supervised discretization: No</td>
</tr>
<tr>
<td>LibSVM</td>
<td>SVM type: C-SVC; Kernel type: Radial basis function: Exp (-gamma*</td>
</tr>
<tr>
<td>ANN-MLP</td>
<td>Hidden layers: 4; Learning rate: 0.3; Momentum applied to the weights during updating: 0.2; Normalize the class to be between –1 and +1; Seed used to initialize the random number generator: 0; Number of epochs to train through: 500; Percentage size of the validation set: 0; Validation threshold to terminate validation test: 20</td>
</tr>
<tr>
<td>RBF</td>
<td>Clustering seed: 1; Minimum standard deviation: 0.1; Number of clusters: 2; Ridge value: 1.0e^-9</td>
</tr>
<tr>
<td>SMO</td>
<td>Kernel: polynomial; Complexity parameter C: 1.0; Epsilon for round-off error: 1.0e^-12</td>
</tr>
<tr>
<td>Random Forest</td>
<td>Maximum tree depth: unlimited; Number of trees: 10; Random number seed: 1</td>
</tr>
<tr>
<td>CART</td>
<td>Heuristic; Minimal number of observations at the terminal nodes: 2; Number of folds in the internal cross-validation: 5; Random number seed: 1; Percentage of the training set size: 1; Pruning decision: 1SE rule minimal cost-complexity pruning: yes</td>
</tr>
</tbody>
</table>
the top three in the tested classification methods to be the best. It is evident from the Figs. 5 and 6 that the ANN–MLP presents the highest classification accuracy of 89.5%, followed by the NB classifier (86.7%). The TP rate is also high for both classifiers (90.2%).

6. Discussion

Babaoglu et al. employed binary particle swarm optimization (BPSO) and genetic algorithm techniques to select useful features from exercise stress testing data.
obtained from 480 patients. Using SVM as a classifier, they obtained 81.46% accuracy (for the BPSO technique) in CAD detection. In another study by the same group, the authors used PCA for dimension reduction on the same dataset and obtained an accuracy of 79.71% with the SVM classifier. However, as indicated earlier, exercise stress testing has its own limitations and hence, obtaining a good representative dataset using this test is difficult.

Karimi et al. presented a technique using wavelet analysis and ANNs for analyzing heart sounds to detect CAD. They detected normal classes with 90% accuracy and CAD classes with 85% accuracy. Arafat et al. studied the use of combined uncertainty methods (fuzzy and probabilistic) in the diagnosis of CAD using ECG stress signals, and they observed an accuracy of around 80%. Zhao and Ma proposed a technique based on the empirical mode decomposition-teager energy operator for feature extraction and back propagation neural network as the classifier to detect CAD from heart murmur signals, and they achieved an accuracy of roughly 85%.

Several studies have also been conducted to evaluate the efficiency of nonlinear features in detecting CAD. The normal fractal properties of RR interval dynamics that were analyzed using DFA were found to be decreased in CAD patients. Such decreased values were also observed in our study (α and D have low values in the case of CAD patients, as indicated in Table 1). A study by Antanavicius et al. on the nonlinear dynamics analysis of ECG signals indicated that recurrences percentage, mutual information, fractal dimension, and embedding dimension error were good quantitative descriptors of fluctuations that were significantly different in 108 CAD patients and 54 patients without CAD. Nonlinear features such as the scaling exponent, correlation dimension, entropy measures, and various prediction/approximation methods were proposed to differentiate between patients with CAD and healthy subjects in addition to the traditional measures of HRV. Some of these features that were extracted in our work were also found to be significantly different (Table 1).

The ECG signals obtained from three recumbent positions, the supine, left lateral, and right lateral positions have been studied. In one such study, Kim et al. developed a multi-parametric measure for classifying normal subjects (20 cases) from patients suffering from two types of CAD (64 cases) — angina pectoris (AP) and acute coronary syndrome (ACS). The measure was constructed using multiple discriminant analysis of several linear and nonlinear parameters extracted from the HRV signal. The authors obtained accuracies of 75.0%, 72.5%, and 84.6% for the control, AP, and ACS groups, respectively. In a similar study, Lee et al. used several linear and nonlinear HRV parameters in classifiers to detect CAD. In a test group of 99 CAD patients and 94 normal patients, the authors found that the SVM classifier presented the highest accuracy (almost 90%) using features from all three recumbent positions. In a similar study performed by the same group, the highest classification accuracy (88.33%) was obtained using the linear and nonlinear HRV features (extracted from the ECG signals obtained in the three recumbent positions) and the SVM classifier. Here, the authors classified normal cases from AP and ACS
categories, unlike in their earlier study, in which they had only classified normal and CAD groups. In 2008, this group included the carotid arterial wall thickness in addition to HRV features and observed a classification accuracy of around 85% to 90% for the SVM classifier.

A literature study, such as the one summarized above, indicated that (1) the highest accuracy of almost 90% was obtained using techniques that used the ECG signals obtained in three recumbent positions, thereby increasing the processing time and energy, (2) some techniques used the exercise stress testing data, which cannot be obtained from patients of lower level of physical fitness, and (3) it is necessary to develop a technique that uses the general recorded ECG signals and still presents equal or higher accuracies than the existing methods. Therefore, we developed a technique that utilized the general ECG signals, extracted HRV based nonlinear features, and used the significant features to evaluate several classifiers. We obtained 89.5% accuracy for detecting CAD using the simple MLP classifier. This accuracy is almost equal to the highest accuracy recorded in the literature. In addition, the proposed technique overcomes the dataset limitations indicated in points (1) and (2) above. Thus, the proposed technique demonstrates the powerful capability of the nonlinear features and the simple PCA technique in differentiating normal from CAD patients. Our future work would include improving our current methodology to determine features and classifiers that can differentiate various sub-groups of CAD, namely, angina pectoris (AP) and acute coronary syndrome (ACS).

7. Conclusions

CAD is a serious illness that has a high mortality rate and does not present itself at its early stages. Therefore, simple, cost-effective, non-invasive, and efficient diagnostic techniques can be used to detect this disease in its early stages in a wide population. Kurt et al. used several demographic and medical history-based parameters in several classifiers to detect the presence of CAD. On analyzing their technique in 1,245 subjects, they obtained an accuracy of 79.1% and an AUC value of 0.783 for the multilayer perceptron classifier. Other studies on CAD detection are based on applying data-mining techniques on various heart-related signals. In this paper, a technique based on HRV signals derived from ECG signals was presented. ECG acquisition is a commonly available technique, and calculating the HRV signals is fast and simple. From these HRV signals, we extracted simple nonlinear features and subjected them to PCA, which is another easily implementable statistical method that can determine the projections, as in this work. We observed that when these projections are used in a commonly available multilayer perceptron classifier, a classification accuracy that is higher than those observed in published CAD studies was obtained. We feel that the technique is of clinical significance as it can be easily implemented and provides good classification accuracy. Future work may include researching other features for improving the accuracy and conducting studies on a wider range of the population.
Acknowledgements

This research was supported in part by U.S. National Science Foundation (NSF) award numbers EPS-0701491 and EPS-1003897.

References


