Circuits and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ens.ewi.tudelft.nl/

CAS-2018-11

M.Sc. Thesis

An automated ECG signal quality assessment method with supervised learning algorithm

Yuyang Wang

Abstract

Wearable health has become a striking area in our daily life. Electrocardiogram (ECG) is one of the biomedical signals collected by the wearable or portable devices, which is widely used in heart rate monitoring and cardiac diagnosis. However, automatic ECG signal analysis is difficult in real application because the signals are easy to be contaminated by the noise and artifacts. Thus, the quality of ECG signals is essential for the accurate analysis. The objective of this project is to design a reliable automated ECG signal quality indicator based on the supervised learning algorithm, which intends to estimate the quality of the signals and distinguish them.

The methodology of this project is creating a classification model to indicate the quality of ECG signals based on the machine learning algorithm. The model is trained by the extracted features based on the Fourier transform, Wavelet transform, Autocorrelation function and Principal component analysis of ECG signals. Subsequently, the feature selection techniques are proposed to remove the irrelevant and redundant features and then the selected features are fed to classification algorithms. The classifier was then trained and tested on the expert-labeled data from the collected ECG signals. Particularly, we focus on the performance of classifier and use the best training model to predict the quality of new ECG signals.

An automated ECG signal quality assessment method with supervised learning algorithm

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Signal and systems

by

Yuyang Wang
born in Shan dong, China

This work was performed in:

Circuits and Systems Group
Department of Electrical Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Delft University of Technology
Department of
Electrical Engineering

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"An automated ECG signal quality assessment method with supervised learning algorithm"** by **Yuyang Wang** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 22-11-2018

Chairman: _____
Prof.dr.ir. A.J. van der Veen

Advisor: _____
Dr. Ir. Richard Hendriks

Committee Members: _____
Dr. Ir. Richard Heusdens

_____
Dr. Ir. Richard Hendriks

_____
Dr. David Tax

Abstract

Wearable health has become a striking area in our daily life. Electrocardiogram (ECG) is one of the biomedical signals collected by the wearable or portable devices, which is widely used in heart rate monitoring and cardiac diagnosis. However, automatic ECG signal analysis is difficult in real application because the signals are easy to be contaminated by the noise and artifacts. Thus, the quality of ECG signals is essential for the accurate analysis. The objective of this project is to design a reliable automated ECG signal quality indicator based on the supervised learning algorithm, which intends to estimate the quality of the signals and distinguish them.

The methodology of this project is creating a classification model to indicate the quality of ECG signals based on the machine learning algorithm. The model is trained by the extracted features based on the Fourier transform, Wavelet transform, Autocorrelation function and Principal component analysis of ECG signals. Subsequently, the feature selection techniques are proposed to remove the irrelevant and redundant features and then the selected features are fed to classification algorithms. The classifier was then trained and tested on the expert-labeled data from the collected ECG signals. Particularly, we focus on the performance of classifier and use the best training model to predict the quality of new ECG signals.

Acknowledgments

My advisor Dr. Ir. Richard Hendriks and the daily supervisor at IMEC Neide Simoes Capela give me a lot of help during the process of master thesis. I really want to thank them due to their guidance and patience. Durig the period at company, my colleagues xiaolin and muqing are really nice and accompany with me in my daily life and work. After I went back to Delft, I also have a good time with my friends Tao, Yefeng and Daiming. Without them, this would be hard to finish this thesis. At last, I want to thank my parents who just support everything I did and give me useful advice in life.

Yuyang Wang
Delft, The Netherlands
22-11-2018

Contents

List of Figures

List of Tables

1 Introduction

Wearable health has become a striking area in our daily life. The electrocardiogram (ECG) is one of the biomedical signals collected by wearable or portable devices, which is widely used in heart rate monitoring and cardiac diagnosis. It records the electrical signal by measuring the voltage difference between the electrodes placed on the skin. The heart rate is a typical feature extracted from the ECG signals. The ECG is a non-stationary and time-varying signal where the intervals of adjacent heart beats vary along with time. A normal cycle of the ECG signal represents a single heartbeat, which always consists of the peaks and ECG waves. Figure 1.1 shows these identified waves of the ECG signal in a normal cycle. The description of these waves in ECG signals is also illustrated below.

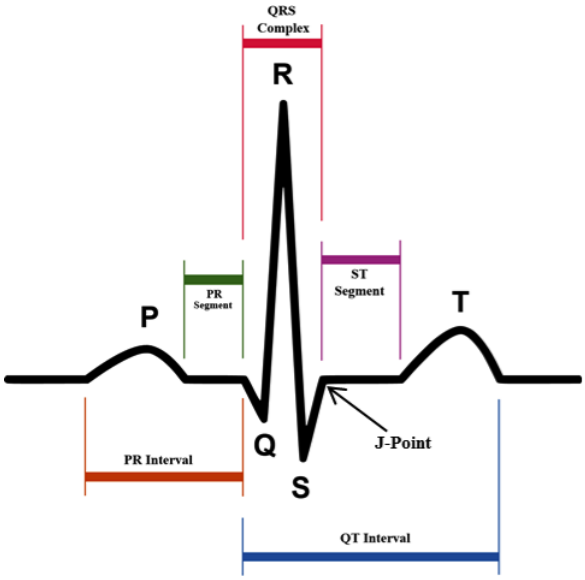


Figure 1.1: The ECG signal

1. **QRS complex**: It is the central and most visually obvious part in the ECG signal, which represents the main voltage variations on the ECG signal. The peak with largest voltage is usually seen as the R wave. The Q wave is the left saddle point near the R peak, and the S wave is the right saddle point near the R peak. The combination of these waves consists of QRS complex.

2. **P wave**: It is the wave with a lower amplitude compared with the R wave which always lies on the left side of QRS complex.

3. **PR interval**: It is a duration measured between the starting point of P wave and the beginning of Q wave.

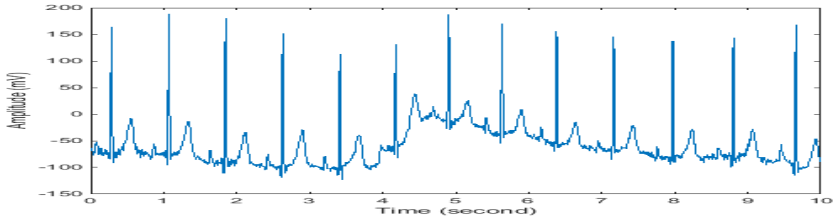
4. **T wave**: It represents the other low amplitude wave which always lies on the right side of QRS complex.

5. **ST segment**: It is a segment that connects the S wave and T wave. [4].

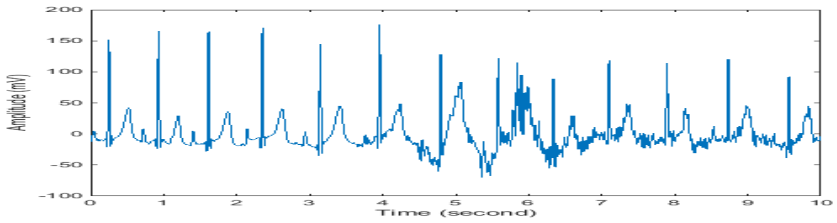
ECG signals can reflect the complex electrical and mechanical processes present in our heart which represents the most valuable information of the heart. Also, ECG is widely used in many applications like heart disease diagnosis, sleep apnea detection and heart rate recording due to its stability and representative of heart activity.

1.1 Thesis Motivation

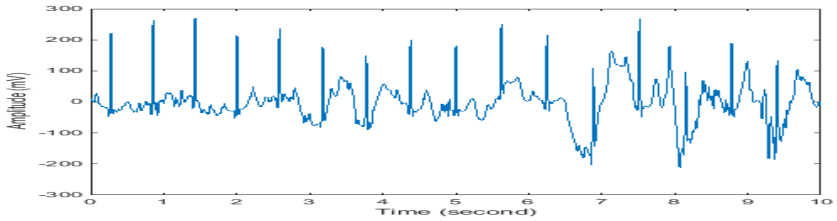
Wearable health has become a striking area in our daily life. Wearable sensors are widely used in health care due to its portability and low cost. The ECG signal is one of the most popular biomedical signals in wearable health. Most applications of ECG signals require noise-free ECG characteristic points and morphological structures for accurate measurements [5]. However, the ECG signals are often contaminated with artifacts such as baseline wander, muscle activity and motion artifacts in the collection phase due to the various activities of the subjects. To visualize the typical artifacts in the ECG signal of wearable devices, Figure 1.2 is shown below.



(a) The ECG signal with baseline wander



(b) The ECG signal with muscle activity



(c) The ECG signal with motion artifacts

Figure 1.2: The typical artifacts in ECG signals

Figure 1.2a illustrates the baseline wander which is a low-frequency noise present in the ECG signal which varies the amplitude of baseline to an extent. Figure 1.2b illustrates the muscle activity which is generated if the subject is nervous or shaky during the collection. Figure 1.2c illustrates the motion artifacts caused by body movements. The body movements usually change the position of electrode of ECG devices which results in the impedance variation between the electrode and the skin. Thus, the output of ECG devices are altered because of the dependence between the voltage and impedance. For the ECG signals with strong motion artifacts, the morphological structure of the ECG signal will be corrupted. In this case, it is demanding to perform accurate analysis of these ECG signals. Thus, automatic assessment of the quality of ECG signals is a crucial step before the measurements in ambulatory wearable recordings.

1.2 Related work

Researchers proposed multiple ECG signal quality assessment methods called SQA before. Most methods aim to identify the collected ECG signals as acceptable or unacceptable. In the SQA methods with linear features and threshold approach, the features of morphological structure and heart beats were extracted from the ECG signals. C.Orphanidou, et al.[6] presented six standard features to estimate the reliable ECG signals. Then the template matching method was presented to make comparison. This method has a decent accuracy but the performance of this method depends on accurate QRS complex detection. In [7], D.Hayn et al. proposed four ECG features based on the amplitude, the crossing point of each lead and the stability of the detector to assess the quality of the ECG signals.

Further, SQA methods based on linear features and machine learning were also presented. These methods were achieved by training the classifier with the extracted features. In [8], G.D.Clifford et al. utilized six signal quality indices as features and then trained them with classifiers like Naive Bayes, linear discriminant analysis, support vector machine and multi-layer perceptron neural network. In [9], J.Kuzilek et al. proposed a three stage algorithm to assess the quality of ECG signals. Statistical features like standard deviation and maximum of ECG signals were extracted and compared with the threshold by the decision rule. Then the decision result computed a score for each observation. Next, new features like kurtosis and covariance matrix were calculated. These features were given to the machine learning algorithm which can produce the score of the result. In the last step, the combination of the first and second score is utilized to identify the quality of ECG signals.

At last, some researchers proposed SQA methods with non-linear features. In these methods, the features represent the temporal and spectral information of ECG signals. J.Lee et al.[10] utilized the empirical mode decomposition (EMD) algorithm for the automatic identification of the artifacts. In addition to features like mean and standard deviation of the first intrinsic mode function (IMF), the non-linear feature shannon entropy is also calculated. In [11], U.Satija presented an automated ECG SQA method with unsupervised learning algorithm based on Wavelet decomposition. The statistical features like maximum absolute amplitude and difference of consecutive dynamic am-

plitude range are computed in the low frequency sub-bands. The the high frequency sub-bands based features are first maximum of autocorrelation function, kurtosis and zero-crossing rate.

The above SQA methods with morphological and RR interval features have some practical limitations illustrated below.

- To extract the QRS features, a reliable QRS complex detector under ambulatory recording environment is needed. It is a challenging task because the noise and artifacts can contaminate the ECG signals and change their morphological structure.
- The performance of the SQA methods will decrease under different ECG morphological structures. These methods need to be tested on more subjects.
- It is challenging to find an optimal threshold to classify the quality of ECG signals. If the ECG signals have overlapping frequency bands of artifacts and ECG waves, this SQA method can provide incorrect decision boundary.

In the EMD based method, the IMFs are calculated in the time domain, which lacks the frequency information. The number of IMFs is demanding to determine due to the variability of ECG structures and artifacts. Apart from that, the EMD algorithm needs expensive computations because of the sifting operations.

In the Wavelet decomposition based method, the ECG signals are decomposed in frequency domain through the convolution with wavelets. It solves the problem that was present in the EMD based method. However, the proposed features extracted from low frequency and high frequency sub-bands are not enough to distinguish the clean and the contaminated ECG signals. More useful features demand to be extracted and analyzed. Also, the hard decision thresholds for proposed features are not accurate due to the variability of ECG signals.

Due to the presented limitations, the ECG signals require a reliable automated quality assessment method which does not rely on an accurate QRS complex detector and hard threshold approaches. In this case, the features of SQA method need to represent the characteristic of ECG signals with a supervised learning algorithm. Moreover, the improvement of time and frequency domain features is also necessary in automated quality assessment methods.

1.3 Proposed Methodology

The objective of this work is to develop a supervised learning method for automatic ECG signal quality assessment which first computes the features from ECG signals and then uses them in the supervised learning algorithm. To increase the quality assessment resolution, the proposed methods calculate features in each ten-second segment. Also, the supervised learning problem needs labels on each observation, which are manually annotated by our researchers on each segment. Due to large variations of ECG signals on different subjects, normalization is used as the preprocessing approach in this proposed methodology. Since the information of ECG waves, noises and artifacts in time

and frequency domain are very crucial in the feature extraction procedure, we proposed three feature extraction methods.

Discrete wavelet transform (DWT) is a widely used technique for analyzing the temporal and spectral information of signals. It is suitable for non-stationary signals with varying frequencies. The DWT can decompose the signal into sub-bands. The sub-bands help to identify the clean and contaminated ECG signals at their corresponding frequencies. Because of this advantage, we use the DWT algorithm as our first feature extraction method.

Our second feature extraction method is based on Autocorrelation function. The periodic information and correlations between different patterns of ECG signals can be extracted from the Autocorrelation function to distinguish the clean and contaminated signals.

Principal component analysis (PCA) is utilized as the last feature extraction method. In this method, we construct a matrix including each QRS complex presented in ECG signals. Then we compute the eigenvalues of this matrix and calculate the first three largest ratios between the eigenvalue and the sum of all the eigenvalues as features by the PCA algorithm. The eigenvalue ratios represent the similarity between the successive QRS complex of ECG signals. The advantage of this method is that it does not demand an accurate R peaks detector. If there are R peaks wrongly detected, it illustrates that there are noise or artifacts in the ECG signals which can also be reflected in the features.

After the feature extraction, the next procedure of this work is feature selection. We use two feature selection methods named filter and wrapper method. The first one is based on attributes of the feature itself, and the latter one is based on a specific classifier. Afterwards, the performance of these two methods are compared and analyzed. Then the feature matrix is constructed based on the selected features of all the observations.

The last procedure of this work is to train a classifier with the supervised learning algorithm. The feature matrix is split into training and validation data. The training data and corresponding labels are utilized to train the classifier with cross validation method. The best classification model is chosen according to the result of cross validation. Subsequently, the validation data are used to test the performance of the model. Finally, the ECG signal quality can be predicted based on the posterior probability produced by the SVM classifier.

The whole scheme of this work consists of five steps which is shown in Figure 1.3.

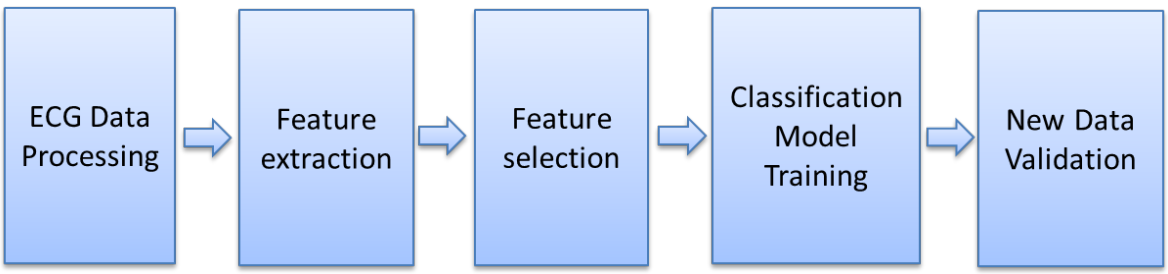


Figure 1.3: The whole scheme of ECG signal quality assessment

1.4 Research Question

Our research aims at designing an automated ECG signal quality indicator with an easily implementable and accurate method. Our proposed approach includes feature extraction, feature selection, model training and classification. The first step includes the methodology about extracting features which focus on capturing the main characteristics of clean (and contaminated) ECG signals. In accordance to the objective of this step, the following research questions are

1. What is the difference between clean and contaminated ECG signals?
2. Which type of features can be extracted from ECG signals?
3. Which method can be used to extract features?

The second step aims at removing the irrelevant and redundant features and choosing best feature subset to classify the clean and contaminated ECG signals. The main research questions of this step are

1. How to identify the importance of each feature?
2. How to choose the best feature subset?

The last step involves training the classification model and identifying the ECG signal quality. The trained classification model can automatically identify the quality of new ECG signals. The main research questions behind this step are

1. Which classifier should be used for this problem?
2. How to evaluate the performance of classification result?
3. How to identify the quality of ECG signals automatically?

1.5 Thesis Contribution

The thesis presents an automated ECG signal quality indicator based on the supervised learning algorithm. The major contributions of the thesis are summarized below.

1. New features are generated based on the Discrete wavelet transform algorithm. In general, the wavelet transform is a favorite technique to extract features of ECG signals due to its ability to capture the temporal and spectral information. In our case, the information of ECG signals in the time and frequency domain are both important. Excepting the traditional statistical features like mean, standard deviation and variance, the new features based on the wavelet transform like Median absolute deviation and Approximate Entropy are proposed in this work.
2. New features are generated based on the Autocorrelation function of ECG signals. The ECG is a non-stationary and periodic time series signal. Because of that, features based on Autocorrelation function is proposed as its ability to describe the periodic information of time series signal. In this work, the features including

periodic information such as the location and amplitude of local maximum peaks, first zero-point location and zero-crossing rate are extracted from Autocorrelation functions.

3. It is proposed to use the Principal component analysis to extract features. The Principal component analysis (PCA) is a mathematical tool for finding patterns in time series signals. To represent similarity between the successive QRS complexes of ECG signals, the features based on PCA algorithm were extracted by calculating the ratio of the first three eigenvalues after sorting the values in descending order over the sum of all the calculated eigenvalues. The main advantage of this method are shown as follows: 1) The value of eigenvalue ratios can reflect the differences in each QRS complex; even if the differences are small; 2) This method does not rely on an accurate R peak detection algorithm.

The objective of this work is to design an automated ECG signal quality indicator, and the classifier we used in this work is supervised SVM which can produce a posterior probability on each observation. Based on that, we propose a method to divide the prediction result into several quality degrees according to the posterior probability.

1.6 Organization of the thesis

The thesis comprises six chapters and can be organized as follows. Chapter 1 presents the introduction that consists of the background, related work, proposed methodology, thesis motivation and thesis contributions. Chapter 2 introduces the wearable devices, the methodology for collecting ECG signals, different types of noises contaminated in ECG signals and the corresponding quality annotation works. Chapter 3 describes the feature extraction methods based on Wavelet transform, Autocorrelation function and Principal component analysis. Chapter 4 illustrates the popular feature selection methods and their corresponding advantages and disadvantages. The implementation of these methods in this work are discussed and the final feature selection result of each method is compared in this chapter. Chapter 5 introduces the utilized classifier to estimate the quality of ECG signals. The performance of classifier on each set of features was also presented in this chapter. Chapter 6, the final chapter, provides a discussion on the proposed methods. In addition, conclusions, limitations and further research are also discussed in this chapter.

2 Data collection and annotation

This chapter elaborates the introduction of the wearable devices, the methodology for collecting ECG signals and how the signal quality annotation works.

2.1 Wearable devices

Wearable or portable devices are smart electronic devices which can be implanted into clothing or worn on the body. It has developed in recent years and already entered the area of healthcare. The wearable devices can be utilized in various biomedical applications like heart rate monitoring, heart condition detection and respiration extraction. Some wearable devices in the form of implants or accessories have already been produced and applied in our daily life such as smart watches, armbands, and glasses [12].

In our research, the ECG signals were collected from one wearable device attached to two different body interfaces of IMEC. The first interface attached to our body is a chest patch for collecting the ECG and acceleration signal. The wearable device was designed to record two channel ECG signals with a sampling frequency 256 Hz and the acceleration signal at 32 Hz. However, the acceleration was not considered in this work. The other interface is a belt with the same device to collect ECG and acceleration signals from different position of our body.

2.2 Data Acquisition

In our experiment, two datasets were collected from the wearable devices. The experimental protocol was different for collecting these two datasets. For collecting dataset one, the ECG belt was used, and six subjects participated in this experiment. The participants were informed to accomplish the measurement under controlled conditions, where the subjects need to perform six consecutive activities as shown in Table 2.1.

Controlled Experiment	
Activities	Time
Sitting	5 minutes
Lying	5 minutes
Walking	2 minutes
Walking upstairs	30 seconds
Walking downstairs	30 seconds
Jumping	30 seconds

Table 2.1: Controlled experiment

After performing the last activity, the subjects need to tap the sensor inside the device to have a marker of the end in the signal and return the device to the researcher.

For collecting dataset two, the ECG patch wearable device was used. In this experiment, four subjects were asked to accomplish the measurement under uncontrolled conditions, where the subjects were not restricted to specific activities. In other words, the subjects were free to perform the activities as usual. Each participant was required to wear the device for around eight hours.

The total duration of the data used in this study was 15 hours. Specifically, the duration was one hour for the controlled experiment and 14 hours for the uncontrolled experiment.

2.3 Quality annotation for ECG signals

To indicate the quality of ECG signals, it is usually in machine learning to divide the raw signals into several small segments using sliding windows. The length of window should be chosen in accordance to the application. Feature extraction on each window can be achieved with low computational complexity by decreasing the window length. However, the segments might not contain enough information to represent the features in ECG signals. In contrast, increasing the length of the window could detect the low frequency ECG morphological structures with the price of increasing the computational time. Thus, the selection of the window length is a trade-off. According to the reference[6][13][8], the commonly used window length is ten seconds, which was the same used in this work.

The annotation work was performed on each segment of single channel ECG signal cut by sliding windows by four annotators that received previous training. The annotators inspected the recording ECG signals and manually labeled the ECG segments into two signal quality levels. The annotation rules of signal quality was summarized as follows.

- Label 1: ECG signals are perfect, have clear QRS complex, P and T waves. There are no obvious noise and abrupt variations in the signal.
- Label 0: ECG signals are not perfect. Baseline wander, transient high amplitude impulse, motion artifacts and muscle activity noise exists in ECG signal.

If the annotators were not sure about the annotations on some segments, all the annotators had to discuss about it and acquired a common result. The annotations performed on 15 hours data from both the controlled and uncontrolled experiment. After the labeling work, the total number of observations of label 1 and label 0 are 2908 and 3908 respectively.

Preprocessing and Feature Extraction of ECG Signal

3

This chapter describes the preprocessing of the original ECG signals and the feature extraction methods based on specific algorithms, which is a necessary step in designing an automated ECG signal quality indicator. The extracted features are essential in the classification problem because it includes the important information of the data. In this work, different types of features representing the ECG signal are proposed, which involves the main information of ECG signals.

3.1 Theoretical Background

3.1.1 Discrete Wavelet Transform

The Fourier transform focus on the spectral information of the signal. In contrast with the Fourier transform, the main advantage of the Wavelet transform is that it captures both the temporal and spectral information. In numerical analysis, the Discrete wavelet transform (DWT) is achieved with the discretely sampled wavelets which has similarity with Discrete fourier transform (DFT). For example, the DWT of a signal $x(t)$ is defined as follows.

$$DWT(i,j) = \frac{1}{\sqrt{2^i}} \int_{-\infty}^{\infty} x(t)\psi(\frac{t-2^i j}{2^i})dt \tag{3.1}$$

where ψ is the wavelet function chosen from the wavelet family, i represents the decomposition level and j is the shifting parameter in DWT. The DWT of signal $x(n)$ is achieved by passing it through the low pass and high pass filters. At the beginning, the signals are convoluted with a low pass filter defined as g. This procedure is illustrated as follows.

$$y(n) = x(n) * g(n) = \sum_{k=-\infty}^{\infty} x(k)g(n-k) \tag{3.2}$$

Simultaneously, the original signal is also convoluted with a high pass filter defined as h. The outputs of the high pass and low pass filter are defined as the detail coefficients and approximation coefficients respectively. The cut-off frequency of these two filters is in the middle of the maximum frequency of the input signals. Thus, half of the frequencies of the input signal are remained after passing through the filter. In this case, only half of the samples should be left according to the Nyquist rule. This process is achieved by downsampling the output signals of the filters. Then the first level decomposition is finished. After that, the next level decomposition is the repetition of the previous step but only for the approximation coefficients. This further decomposition of the approximation coefficients is processed by passing it through the filters and then downsampling again. The advantage of DWT in each level decomposition is that the frequency resolution of the output signal is doubled because only half the

frequencies of the input signal have remained. However, the temporal resolution of the original signal is reduced due to the downsampling. This procedure will continue until multi-level decomposition is finished. For instance, the diagram of DWT with third level decomposition is illustrated in Figure 3.1.

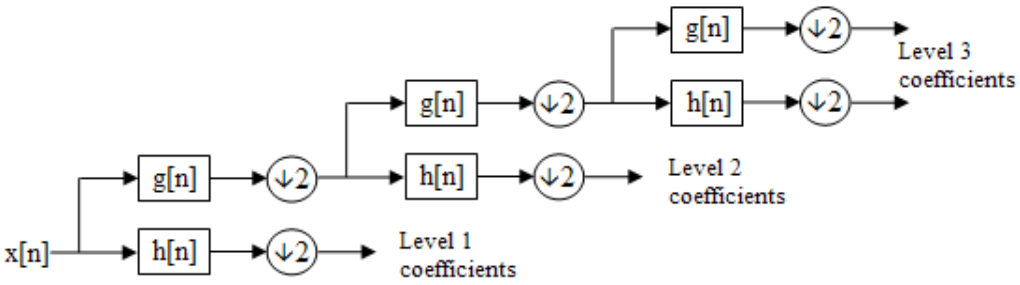


Figure 3.1: The diagram of third level discrete wavelet transform

In each level wavelet decomposition, the input signal is divided into the detail and approximation coefficients. From equation 3.1, we can see that the number of samples in the original signal has to be a multiple of 2^i where i indicates the decomposition levels. For the signal with 64 samples and the maximum frequency fm, the frequencies and samples of it in fourth level decomposition are shown at Table 3.1.

Level	Frequency	Samples
4	0 to $f_m/16$	4
4	$f_m/16$ to $f_m/8$	4
3	$f_m/8$ to $f_m/4$	8
2	$f_m/4$ to $f_m/2$	16
1	$f_m/2$ to f_m	32

Table 3.1: The frequency range and samples in signal decomposition

From table 3.1, we can clearly see that the frequency range and samples of coefficients are halved in each decomposition.

3.1.2 Autocorrelation function

Autocorrelation (ACF) is a measure of the similarity between different patterns of a signal. The correlation is achieved by multiplying the original signal with a shifted copy of itself. The main objective of the ACF is finding repeating observations of the signal. It is commonly used in analyzing the periodic information of the time series signal. In statistics, the normalized ACF of a signal $x(n)$ is computed in equation 3.3.

$$R(k) = \frac{E[(x_n - \mu)(x_{n+k} - \mu)]}{\sigma^2} \tag{3.3}$$

where μ is the mean value and σ^2 is the variance of $x(n)$ that aims to normalize the ACF. The advantage of the normalized ACF is that it reduces the influence of differences on

the amplitude of signals. The maximum value of the normalized ACF is 1. We can see that the ACF is a function of the shift delay k with correlations between different values of the signal. The value of the ACF should lie in the interval between -1 and 1 from anti-correlation to high correlation, where 0 shows that no correlation between two observations. For a discrete time series signal $x(t)$ with n observations $(x_1, x_2, ..., x_n)$, an estimation of the standardized ACF is obtained as

$$R(k) = \frac{1}{(n-k)\sigma^2} \sum_{t=1}^{n-k} (x_t - \mu)(x_{t+k} - \mu) \quad , 0 < k < n \tag{3.4}$$

where the mean of the signal is μ and the variance is σ^2. In conclusion, the Autocorrelation function is a method to describe the correlation between the past observations and the future observations of the signal which depends on the delay they are separated by. It is a useful tool to analyze the periodic information in the time series signal.

3.1.3 Principal Component Analysis

In signal processing and data analysis, the Principal component analysis (PCA) is a mathematical method based on the eigenvalue decomposition of the covariance matrix that reduces the dimensions of the data and remains the most important information of the original data. The first principal component of this algorithm gives a record of the maximum variance in the data. The subsequent principal component has the maximum variance in the remaining data. Besides, the matrix in the PCA algorithm usually needs to subtract the mean value of it to center the data before the eigenvalue decomposition. After that, eigenvalue decomposition of the computed covariance matrix is implemented. The diagonal eigenvalue matrix and its corresponding eigenvectors will be calculated then [14]. The singular value decomposition (SVD) is an essential matrix decomposition method in signal processing and statistics. For example, the SVD of a matrix $X_{m \times n}$ is shown below.

$$X = U \Sigma W^T \tag{3.5}$$

where Σ is a singular value diagonal matrix; U represents the left singular matrix which is a square matrix of order m; W is the right singular square matrix with order n. In both matrix U and W, the columns are unit vectors which are orthogonal to each other. Simultaneously, the SVD of the computed covariance matrix can be written as

$$X^T X = W \Sigma^T U^T U \Sigma W^T = W \Sigma^2 W^T \tag{3.6}$$

where the eigenvectors matrix of $X^T X$ are equal to the right singular matrix W of matrix X and the singular values of matrix X are equivalent to the square-root of the eigenvalues in $X^T X$. The eigenvectors matrix W can be seen as the directions of the projection and the eigenvalues can be considered as variance in this direction of projection. After the matrix decomposition, we need to sort the eigenvectors matrix W with its corresponding eigenvalues in descending order. Then our original matrix X needs to multiply with the sorted eigenvectors matrix W. This process will transform our data matrix into a principal component matrix with each column representing the variance of each direction. The transformation of the original matrix is written as

$$T = XW = U \Sigma W^T W = U \Sigma \tag{3.7}$$

Thus, the left columns of the transformation matrix T corresponding to high singular value represent the most principal components in the signal. In addition, this transformation can also reduce the dimension of our original matrix, which is achieved by keeping only the first N principal components with multiplication of the first N eigenvectors. This process is illustrated as

$$T_N = XW_L = U\Sigma W^T W_L = U_L \Sigma_L \tag{3.8}$$

where the matrix T_N has only N columns now and achieves the dimensional reduction. Therefore, the PCA algorithm is a functional method for visualizing the main components and reducing the dimensions of the datasets.

In summary, the PCA algorithm includes the computation of the covariance matrix, sorting the eigenvalues in descending order with its corresponding eigenvectors and then projecting the data matrix by multiplying with the sorted eigenvectors. The most variability of the original data are involved in the first few projections.

3.1.4 Approximate entropy

In signal processing, Approximate Entropy (ApEn) is a non-linear parameter to illustrate the regularity and unpredictability of variations in the signal. The ApEn quantifies the similarity between the succeeding patterns of the signal. The signal that includes many repetitive patterns with high regularity will produce a small value of ApEn, while the less predictable signal will generate a large value of that [15].

The algorithm for computing ApEn is shown as follows. For a time series signal $x(t)$ with N samples. Before calculating the ApEn, two parameters pattern length k and tolerance of similarity r need to be chosen. Then the whole algorithm of ApEn can be summarized as

1. Generate $k - vectors$:

$$X(m) = [x(m), ..., x(m+k-1)],\ 1 < m < N - k + 1 \tag{3.9}$$

2. Calculate the distance between two vectors $X(m)$ and $X(n)$ by finding the maximum absolute value of difference between them:

$$d[X(m), X(n)] = max[|X(m) - X(n)|] \tag{3.10}$$

3. Define a parameter $C_r^k(m)$ for each $m = 1, ..., N - k + 1$, which is

$$C_r^k(m) = N_k(m)/(N - k + 1) \tag{3.11}$$

   where $N_k(m)$ is the number of $d[X(m), X(n)]$ less than the tolerant parameter r.

4. Logarithmic the $C_r^k(m)$ parameter and average it.

$$\phi^k(r) = \frac{1}{N-k+1} \sum_{m=1}^{N-k+1} ln(C_r^k(m)) \tag{3.12}$$

5. Calculate $\phi^{k+1}(r)$ for k+1 and then the ApEn is defined by

$$ApEn(k, r, N) = \phi^k(r) - \phi^{k+1}(r) \tag{3.13}$$

Thus, $ApEn$ quantifies the difference between the patterns of next intervals and this interval. The small value of $ApEn$ indicates that there is a high similarity between the succeeding patterns of observations. The high value of $ApEn$ implies that the previous observations can not be predicted by the additional measurements, which indicates the time series signal is highly irregular.

3.2 Preprocessing

Preprocessing of raw ECG signal is a crucial procedure before the feature extraction, which involves two steps:

1. Band-pass filtering: the strong artifacts during the collecting phase could influence the quality of ECG signals such as motion artifact, muscle activity and baseline wander. To remove the strong noise and artifacts, the original ECG signals were filtered by a zero-phase, third order Butterworth band pass filter with the cut-off frequency at 0.5 and 40 Hz [16]. This filter helps to remove the low frequency and high frequency noises while does not change the main structure of ECG signals.

2. Normalization: it normalizes the amplitude of the filtered signal from 0 to 1 which reduce the influence of different amplitude across sensors. Let y be the filtered signal, the normalized signal y_{norm} is determined as

$$y_{norm} = \frac{y - min(y)}{max(y) - min(y)} \tag{3.14}$$

In addition, the raw ECG signals were divided into several small segments using sliding windows as mentioned in Chapter 2. The length of window is selected as ten seconds with a trade-off of the computational cost and the information involved in the signal. Then the relevant features will be extracted from each window of signals. To increase the number and variability of observations in the ECG belt data sets, an overlap sliding window was used to divide the ECG signals collected from ECG belt. In this work, 80% overlapping window was used and 1776 segments of ECG signals were extracted with this overlapping window method, compared with 360 segments by non-overlapping window method.

3.3 Feature extraction

Features are used to represent patterns in summarized way including as much information as possible in signals. In this work, the features are extracted from each segment of ECG signals dividing by sliding windows. The features need to train the classification model and will influence the final performance of the quality assessment. Thus, the distributions of selected features on different class should separate very well. The objective of this procedure is extracting representative features to distinguish the clean and contaminated ECG signals in order to obtain a high classification performance.

3.3.1 Wavelet transform based feature extraction method

The original ECG signal can be contaminated by noises and artifacts in the signal acquisition phase. In general, the noise and artifacts will be spread randomly in time and frequency domain of ECG signals. In this case, the features based on DWT algorithm was proposed in this work because of its ability to capture both the temporal and spectral information. The selection of appropriate wavelet and decomposition level is essential in the wavelet transform of ECG signals. In previous work[17] [18], the Daubechies wavelet of sixth order (db6) is chosen from the wavelet family because of its similar structure with the QRS complex. The decomposition level is chosen in accordance to the maximum frequency in the given signals. In preprocessing of original ECG signals, 0.5 to 40 Hz band pass filter was used, which means the maximum frequency in ECG signal is 40 Hz. Thus, six levels wavelet decomposition are taken in this work. After the decomposition, six detail coefficients D1-D6 are extracted from original ECG signals. To illustrate the DWT algorithm on ECG signals, we show the level six wavelet decomposition of the clean and contaminated ECG signal in Figure 3.2, 3.3 and 3.4.

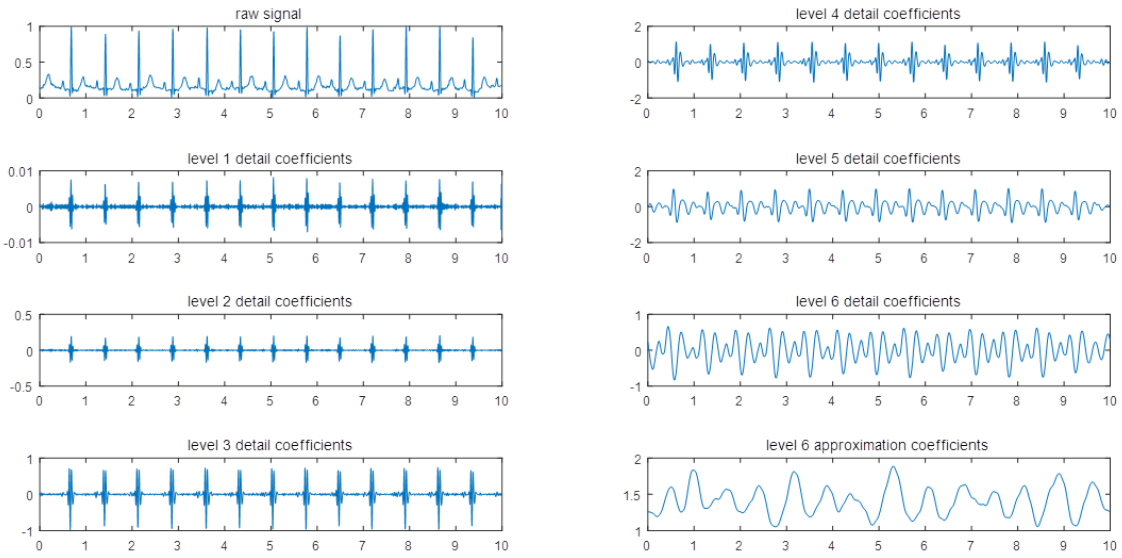


Figure 3.2: The wavelet decomposition of clean ECG signal

The wavelet coefficients of ECG signal with motion artifacts are plotted in Figure 3.3.

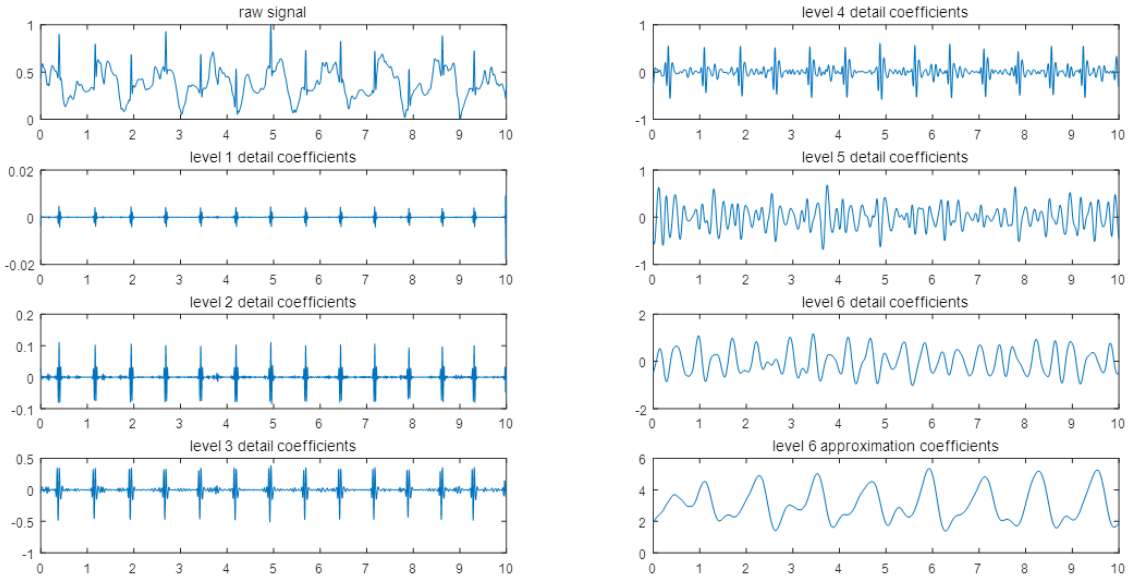


Figure 3.3: The wavelet decomposition of ECG signal with motion artifact

For ECG signal with high frequency noise, the corresponding wavelet coefficients are plotted in Figure 3.4.

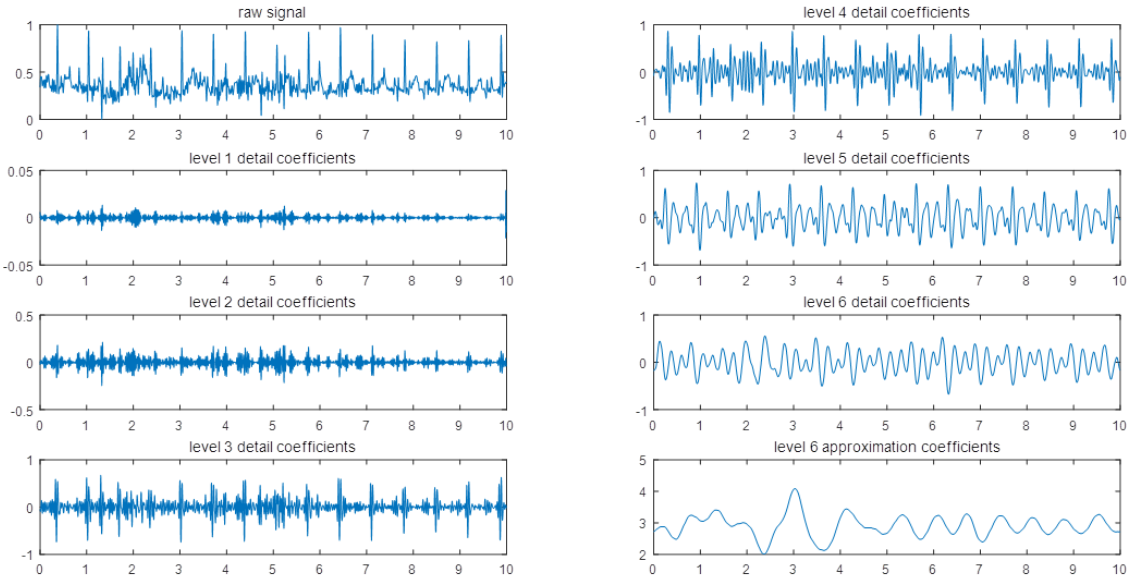


Figure 3.4: The wavelet decomposition of ECG signal with high frequency noise

In Figures above, the left part illustrates the high frequency coefficients of the ECG signal, and the right part shows the low frequency coefficients. We can see that the components remain its shape of original ECG waves because the selected DB6 wavelet has a high cross correlation with the ECG signal. The noise and artifacts of contaminated ECG signals will display at its corresponding frequency coefficients. It can be seen that the clean and contaminated signals differ a lot in specific frequency

sub-bands in accordance to the artifacts. Thus, the extracted wavelet coefficients are useful in ECG signal quality analysis because of its ability to indicate the temporal and spectral distribution of ECG signals.

To describe the difference of wavelet coefficients on the clean and contaminated ECG signal, statistical features over the wavelet coefficients are extracted. The temporal and spectral distribution of ECG signals are represented by the following statistical features: a. mean of the absolute value in each wavelet coefficient; b. standard deviation in each wavelet coefficient; c. median absolute deviation (MAD) in each wavelet coefficient.

An overview of these features extracted from the wavelet coefficients is presented in Figure 3.5.

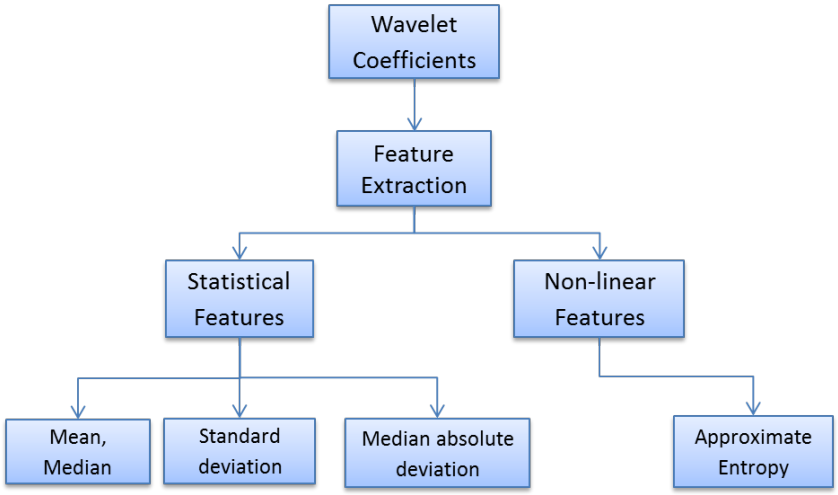


Figure 3.5: An overview of features extracted from wavelet coefficients

In statistics, the MAD is also a statistical parameter that measures the variability of the signal. But the MAD is more functional for outlier detection than the variance. For a time series signal $Y = (y_1, y_2, ..., y_n)$, the MAD is defined by the equation:

$$\text{MAD} = \text{median}(|Y_i - \tilde{Y}|) \tag{3.15}$$

where $\tilde{Y}$ is the median value of Y. The formula starts with the deviations from the median of the signal, the MAD is the median of their absolute values. Apart from these features, the non-linear feature approximate entropy which represents the regularity of patterns in wavelet coefficients was also calculated from the wavelet coefficients. The approximate entropy can be used to distinguish the clean and contaminated ECG signals because of its ability to quantify the complexity of the data. To perform the distribution of the features on different class, the fisher score method was used to evaluate each feature in this work [19]. The principal of fisher score method is illustrated at Appendix definition 2, and the fisher score of each feature is shown in Appendix Table A.1. Finally, the distributions of top three features extracted from DWT algorithm on different class are illustrated in Figure 3.6.

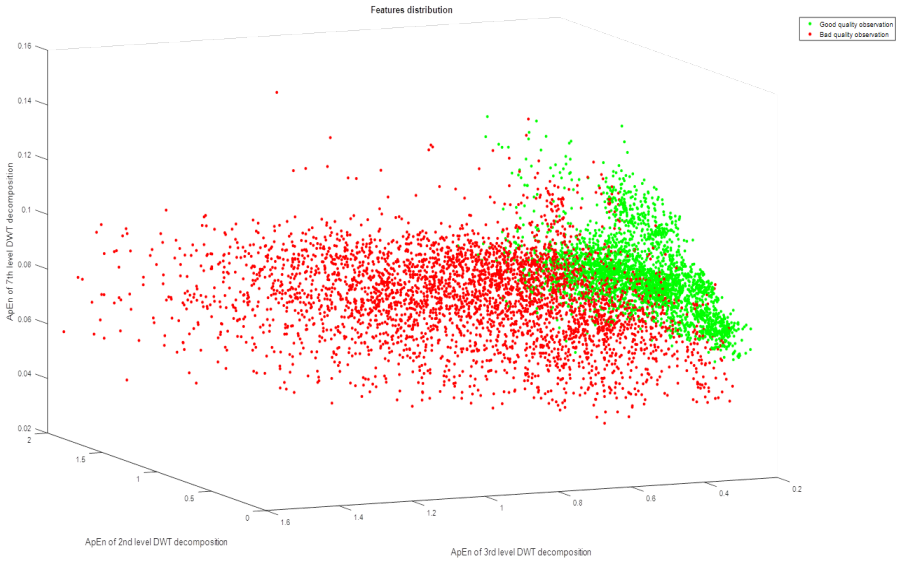
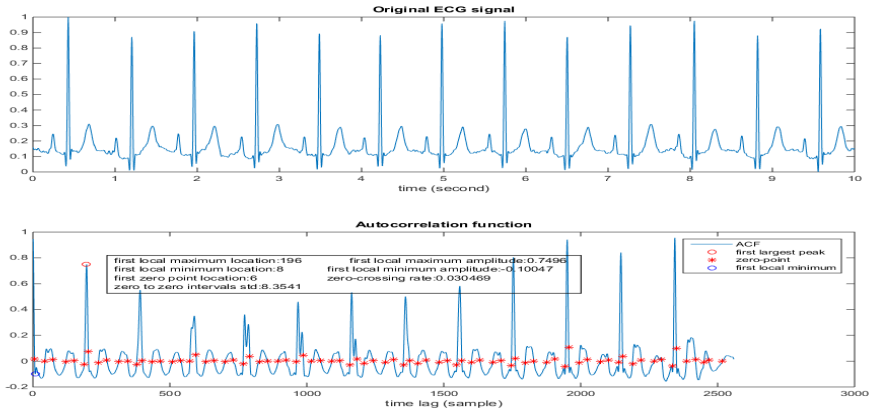


Figure 3.6: The distribution of top three features on DWT algorithm

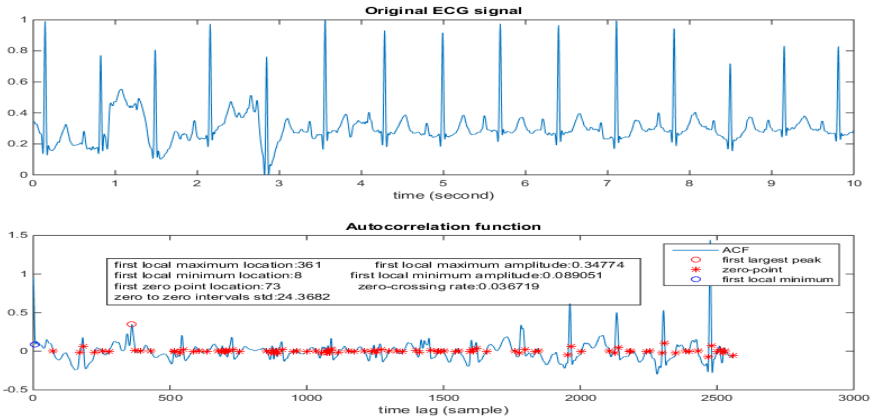
In Figure 3.6, the green colour points represent the good quality observations and the red colour points represent the bad quality observations. It can be seen that the green points are concentrated at the top-right corner of the plot. In contrast, the red points are more likely to locate at the bottom side with more sparse distributions. Thus, it is easy to separate the good and bad quality observations and the final classification result will be illustrated in Chapter 5.

3.3.2 Autocorrelation function based feature extraction method

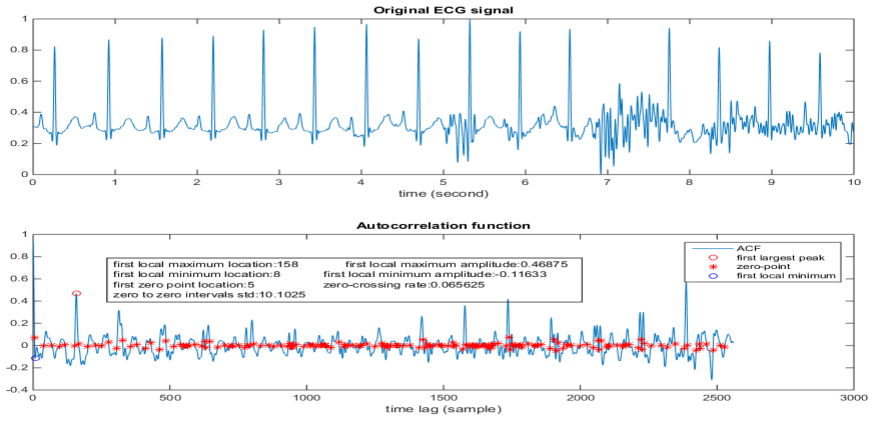
In the theoretical background we mentioned that, the periodic information of ECG signals can be extracted from the Autocorrelation function (ACF). Thus, the features in ACF can possibly separate the clean and contaminated signals due to the significant differences between the ACF of a clean periodic signal and a contaminated signal. In order to demonstrate this statement, the Autocorrelation functions for clean and contaminated ECG signals are illustrated in Figure 3.7.



(a) The Autocorrelation function of the clean ECG signal



(b) The Autocorrelation function of the ECG signal with motion artifacts



(c) The Autocorrelation function of the ECG signal with high frequency noise

Figure 3.7: The Autocorrelation function of clean and contaminated ECG signal

From Figure 3.7, we can see that the Autocorrelation function of clean ECG signal has clear peaks with a decreased amplitude which represents the periodic ECG waves

in the original ECG signal. Besides, there are small up and downs between each pair of peaks which illustrates the information of P wave and T wave. The Autocorrelation function of ECG signal with motion artifacts looks quite different. The periodic peaks are contaminated by the artifacts, and the amplitude of peaks varies a lot. The Autocorrelation function of a contaminated ECG signal with high frequency noise looks similar with the clean ECG signal because the amplitude high frequency variations are small compared with the ECG waves. However, the high frequency noise increases the number of up and downs between each pair of peaks, which can be used to extract features to distinguish them. The maximum time lag is the length of samples in the segment including the autocorrelation coefficients from all the morphological structures in this segment. A number of differences can be observed from the above figures. To characterize the ECG segments, eleven features derived from the ACF are used. The features are described below.

1. The amplitude and location of the first three local maximum in the ACF (except the first point). In general, the local maximum represents the correlation between the heartbeats. The small value of the amplitude indicates two possible artifacts: high- frequency muscle activity between the heartbeats and low-frequency baseline wander caused by electrode movement.

2. First local minimum amplitude and location: the selection of this feature was based on the assumption that a shift of the R-peak towards the deepest point of S-wave represents the first local minimum of the ACF. If the baseline and high frequency noise change the shape of S waves, it can be detected by this feature.

3. First zero-point location: this feature is the interval from perfect correlation to no correlation in the ACF which depends on the morphological structure of ECG signals (QRS complex waves). The noise varying the morphological structure of ECG signals will also be detected in this feature.

4. Zero-crossing rate: this feature is the ratio of zero-crossing numbers over the sample numbers. This feature aims to detect the variations in ECG signals. The larger the zero-crossing rate is, the more likely that the ECG segment contains abrupt alterations.

5. The variance of the zero-point to zero-point intervals: this feature also detects the variations of amplitude in ECG signals. The high value of the variance indicates that significant variations exist in the ECG segment.

The schematic overview of these features based on Autocorrelation function is shown in Figure 3.8.

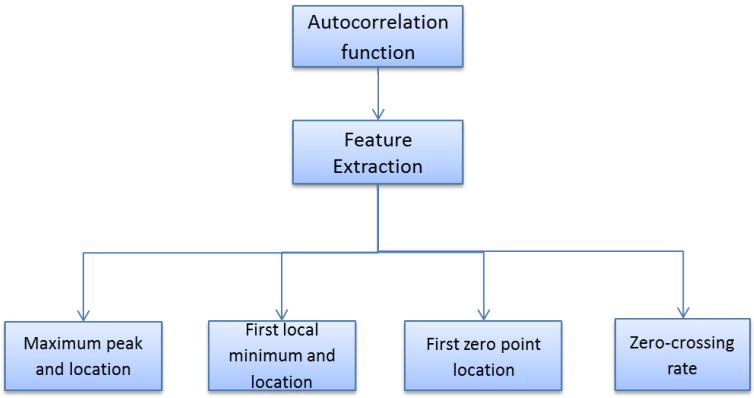


Figure 3.8: An overview of features extracted from Autocorrelation function

The above features based on Autocorrelation function are proposed to separate the clean and contaminated ECG signals. To illustrate how good the feature is, the fisher score method mentioned before was also used to rank the extracted features. The fisher score of each feature is shown in appendix Table A.2. Finally, top three features marked at the table are selected to perform the distributions on different class, which are illustrated in Figure 3.9.

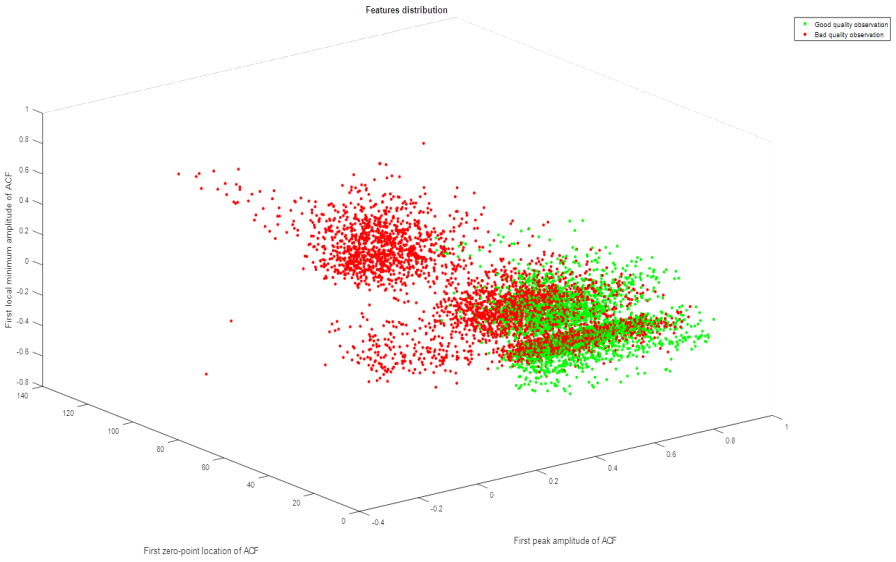


Figure 3.9: The distribution of top three features on ACF algorithm

From the distribution of these features, we can see that some bad quality observations were far away from the good quality observations but it still has some bad quality observations that merge with the good quality observations together in the clusters. In other words, it would be difficult to separate the good and bad quality observations by using a linear hyperplane with a small classification error.

3.3.3 Principal component analysis based feature extraction method

The PCA is a mathematical algorithm for finding periodic information in time series signals. To represent similarity between successive QRS complexes of ECG signals, the features based on PCA algorithm were extracted by calculating the ratio of the first three eigenvalues after sorting the values in descending order over the sum of all the computed eigenvalues. The ratio of eigenvalues quantifies the importance of the principal components in signals. The principal component with low eigenvalue indicates that it contributes little to the variances in the signal. The construction of the objective matrix is a crucial step in feature extraction. The generated objective matrix was based on an ECG heart beat detection algorithm, Pan and Tompkins algorithm[20]. The Pan and Tompkins algorithm is the most widely used algorithm for the QRS complex detection of ECG signals due to its simplicity. In summary, this algorithm consists of the procedures below.

1. Cancellation DC drift and normalization

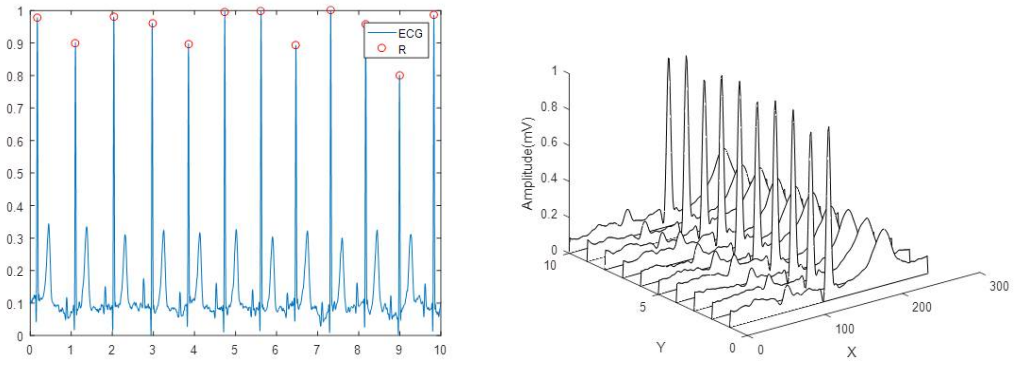
2. Band-pass filtering

3. Derivative filtering

4. Squaring and moving window integration

5. Thresholds adjustment

After the Pan and Tompkins detection algorithm, the locations of R peaks were computed by finding the index of the maximum amplitude of the signals over the threshold. The R peaks location could help us to segment the ECG signal and then construct the objective matrix. The segments were extracted by 0.5 second either side of each R peak. The interval was chosen as 0.5 second because it will contain the main structure of ECG signal (QRS complex, P and T waves). At last, the segments represent the columns in the constructed matrix. As mentioned before, the rows of matrix correspond to observations and columns correspond to variables. Thus, the observations in the constructed matrix are the amplitudes at different time and the variables are consecutive QRS morphological structures in ECG signals. The peak detection result and constructed matrix for a clean ECG signal are shown in Figure 3.10.

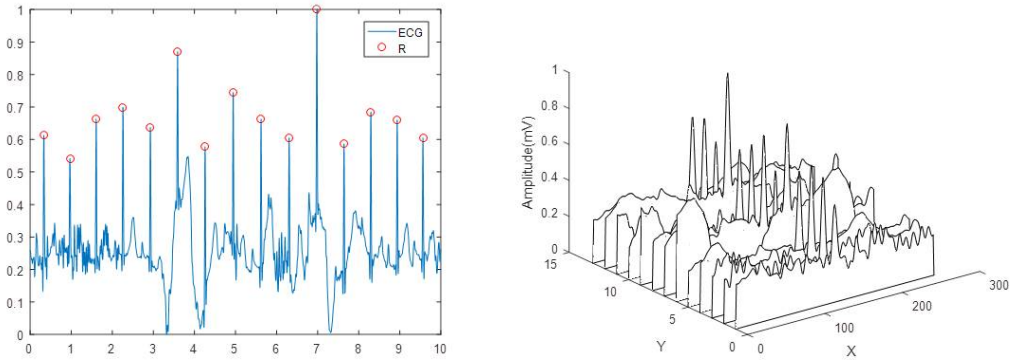


(a) The R peaks detection of a clean ECG signal

(b) The constructed matrix of a clean ECG signal for PCA

Figure 3.10: The R peaks detection result and the corresponding built matrix for a clean ECG signal

For a contaminated ECG signal with motion artifacts, the constructed matrix is shown in Figure 3.11.

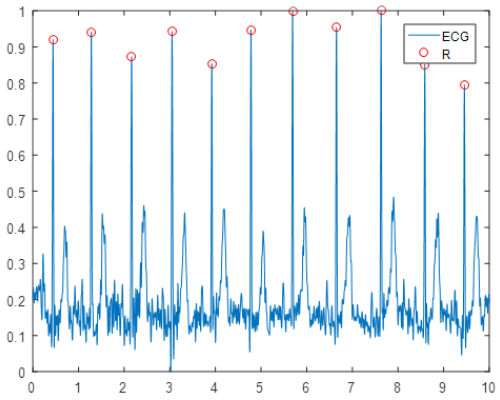


(a) The R peaks detection of a contaminated ECG signal with motion artifacts

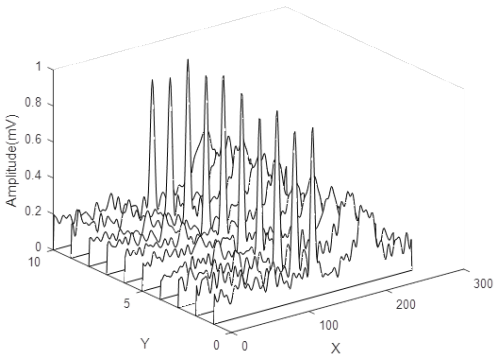
(b) The built matrix for a contaminated ECG signal with motion artifacts

Figure 3.11: The R peaks detection result and the corresponding built matrix for a contaminated ECG signal

At last, the constructed matrix for a contaminated ECG signal with high frequency noise is shown in Figure 3.12.



(a) The R peaks detection of a contaminated ECG signal with high frequency noise



(b) The built matrix for a contaminated ECG signal with high frequency noise

Figure 3.12: The R peaks detection result and the corresponding built matrix for a contaminated ECG signal

After using the PCA algorithm, the principal components and their corresponding eigenvalue ratios are computed. The first two principal components of a clean ECG signal are shown in Figure 3.13.

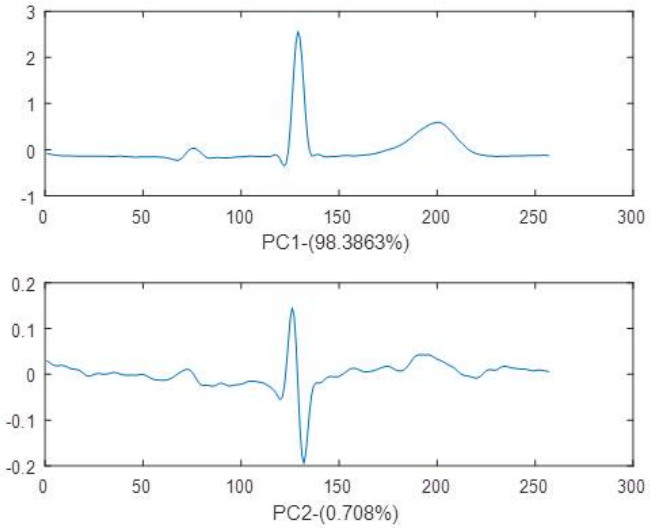
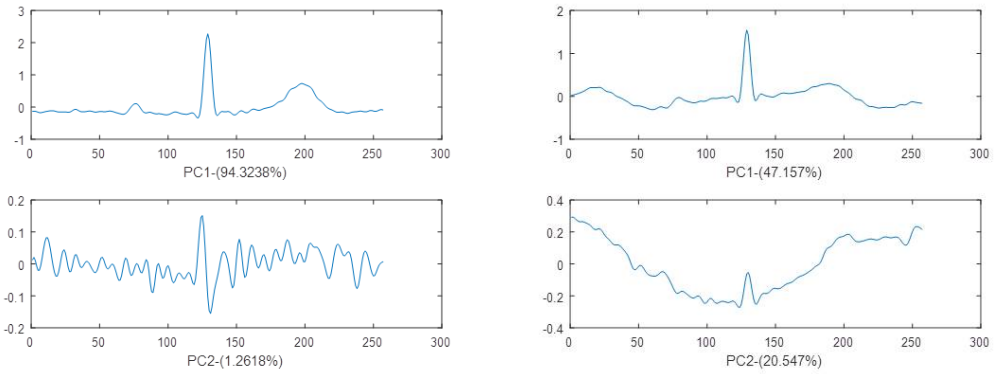


Figure 3.13: The principal components of a clean ECG signal

From Figure 3.13, it can be seen that the eigenvalue ratio of first principal component is over 98% for the clean ECG signals, which occupies almost the whole components of the observations of ECG signals. The reason is that each observation looks very similar in the constructed matrix. There are no abrupt changes of morphological structure in the successive observations. But for the contaminated ECG signals, the corresponding eigenvalues ratios are much different which are illustrated in Figure 3.14.



(a) The principal components of a ECG signal with high frequency noise

(b) The principal components of a ECG signal with motion artifact

Figure 3.14: The principal components of contaminated ECG signals

We find that the eigenvalue ratio of first principal component in the contaminated ECG signal is smaller than it in the clean ECG signal, especially for the ECG signal with motion artifacts. At the same time, the eigenvalue ratio of second principal component in the contaminated ECG signal is higher than it in the clean ECG signal. To illustrate the performance of these features, the fisher score result of each feature is shown in Appendix Table A.3. The distributions of top three features are plotted in Figure 3.15.

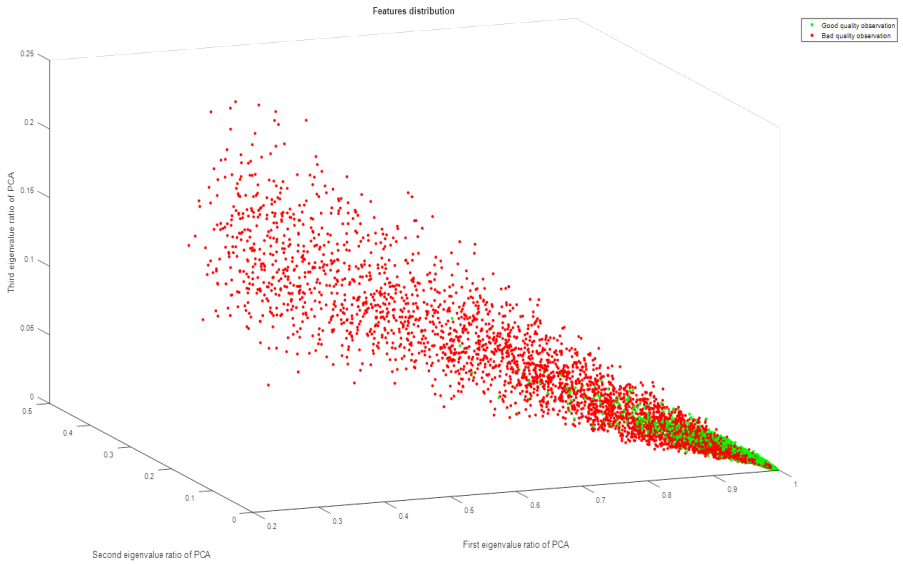


Figure 3.15: The distribution of first three eigenvalue ratios on PCA algorithm

We can see that the shape of the whole distribution looks like a cone and the good quality observations are concentrated at the top corner.

In conclusion, the first principal component represents the main morphological structure in the observations. For the clean ECG signal, all the observations in the matrix

have similar patterns. In contrast, the observations may have different patterns for the contaminated ECG signal because of the noise. Thus, the first component eigenvalue ratio of clean signal is much higher than of the contaminated signal. In addition, the second eigenvalue ratio usually represents the noise in ECG signals. From figure 3.4, we can see that the high frequency and low frequency noises and artifacts are involved in second principal component. This is the reason why the second eigenvalue ratio of contaminated signal is much higher than the clean signal. Thus, the eigenvalue ratios of principal components are chosen as features in view of its difference on clean and contaminated ECG signal.

3.3.4 Normal feature extraction method

Apart from the feature extraction method we mentioned before, we also use the original ECG signal and the Fast Fourier Transform (FFT) of that to extract features. The features of the original ECG signal and the FFT include the temporal and spectral information of the signal respectively. These two feature extraction methods are utilized in this work because of their simplicity, and all the extracted features of these two methods are shown in Table 3.2.

Methods	Features
Original signal	mean, standard deviation, MAD, kurtosis, skewness, ApEn
FFT	mean, max, standard deviation, kurtosis, skewness, ApEn

Table 3.2: The extracted features of the original ECG signal and FFT

The ApEn is the only non-linear feature. All the other features are normal statistical features. The kurtosis is a parameter to measure the outliers of the signal. The signal with the high value of the kurtosis tends to have massive outliers [21]. The definition of kurtosis for the dataset X is

$$kurtosis = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^4/N}{s^4} \tag{3.16}$$

The skewness is a parameter to measure the symmetry of the signal. The skewness of symmetric signals is near zero, and the definition of it is

$$skewness = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^3/N}{s^3} \tag{3.17}$$

where $\bar{X}$ is the mean, s is the standard deviation, and N is the size of the data.

In conclusion, five feature extraction methods based on the original signal, FFT, DWT, ACF and PCA are used in this work. A total of 58 features are extracted from these methods. Before training the classifier, the best set of features need to be selected. The feature selection work is introduced in the next chapter.

Feature selection for ECG quality classification

4

4.1 Introduction

The extracted features are essential in the supervised learning algorithm. The classifier can compute the decision boundary based on the input features to identify the dataset. However, the irrelevant and redundant features might exist in the extracted features. In this case, the performance of the classifier will deteriorate. Feature selection method is proposed to choose the best feature subset from the extracted features and improve the final performance of the classifier. Feature selection is an essential step before the classification, which can also reduce the computational complexity and simplify the classification model. In general, there are three fundamental methods in the feature selection category which are filter, wrapper and hybrid methods[22].

A. Filter methods consider the attributes of feature itself like consistency, similarity and statistical distribution, regardless of the classification model. The filter method can also be classified into univariate filter method and multivariate filter method. Univariate methods usually rank a single feature based on the parameter like symmetrical uncertainty, fisher score and relief. In contrast, the multivariate methods assess an entire feature subset such as MRMR and correlation based methods. The generation of feature subsets depends on features searching methods. In general, there are three usual searching methods: forward feature selection, backward feature reduction and heuristic selection. Forward feature selection is commonly based on adding new features to an empty feature subset. Backward feature reduction is typically based on eliminating features from the whole feature subsets. The heuristic algorithm is utilized in the heuristic selection to explore the generation of the feature subset [23].

B. Wrapper methods is achieved based on the performance of features in the classifier. A large number of classifiers like decision trees, K nearest neighbours and SVM can be used in the wrapper method to asses the performance of the feature subset [24]. The generation of feature subsets also depends on a specific search method. Usually the classification performance of feature subsets obtained by wrapper method was better than filter method in view that the features are assessed based on a real classification algorithm. However, wrapper methods are more computationally expensive than filter methods because of their dependence on the performance of classifiers, especially when using cross-validation.

C. Hybrid methods were achieved by combining the the advantages of filter and wrapper methods. The filter method is used to eliminate the irrelevant features based on the characteristics of features. The wrapper method is then utilized to extract the best feature set. The Hybrid method usually achieves high performance because it combines the best properties of two feature selection methods.

4.2 Methodology

The filter, wrapper, and hybrid methods were all utilized to remove the redundant and irrelevant features. The performance of the feature subset selected by these three methods in classification are compared in Chapter 5. The filter method we used is based on traditional statistic analysis. The utilized wrapper method is a recursive feature elimination algorithm based on SVM classifier.

4.2.1 Wilcoxon signed rank test

In statistics, the Wilcoxon signed rank test is a non-parametric test. The objective of this test is identifying whether two related samples have the equivalent distributions. It does not rely on the assumption of a specific distribution (e.g. normal distribution) of the data unlike t-test. However, it is nearly as efficient as the t-test on samples with normal distributions. In this work, the two tested samples are the features extracted from clean and contaminated ECG signals. To illustrate the dependency of these two samples, we calculate the mutual information between them. If the mutual information is not equal to 0, then these two samples are related. Finally, we find that all the computed mutual information is larger than 0. Thus, these two tested samples are relevant and can be utilized in Wilcoxon signed rank test. For two samples X and Y, the null hypothesis of the two-sample Wilcoxon test is

$$H_0 : \ P(X \ > \ Y) = P(Y \ > \ X) \tag{4.1}$$

Since we are assuming our distributions are equal, what we will try to disprove in this test is that both samples have the same distribution with same median. P value is the crucial parameter in the Wilcoxon rank-sum test which is the probability of the occurrence of the null hypothesis. P value less than the default significance level 0.05 indicates strong evidence to reject the null hypothesis that these two samples have equivalent medians [25]. Figure 4.1 shows the diagram of how to identify the hypothesis.

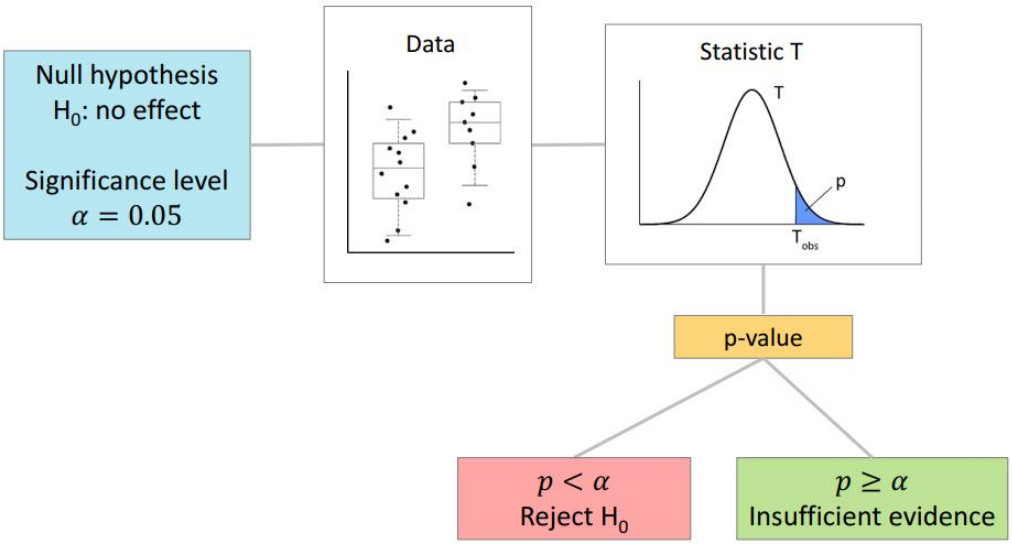


Figure 4.1: The diagram of statistical test [1]

The Wilcoxon test is based on ranking the difference between two samples A and B [26]. Each observation of the combined samples has a rank based on its value. The whole process of this method is shown below.

1. Compute the difference d_i of paired observations from two samples.
2. Rank the absolute value of d_i (give rank 1 to the smallest value).
3. Calculate the sum of the ranks of the positive and negative differences denoted by W^+ and W^- respectively.
4. Select the minimum value of W^+ and W^-, defined by $W = min(W^+, W^-)$.
5. Use tables of critical values for the Wilcoxon signed rank test to find the p-value.

4.2.2 Correlation based feature selection

The Correlation based feature selection (CFS) algorithm is accomplished in accordance to the correlations. The symmetrical uncertainty based on mutual information was used to measure the correlations in this work because of its simplicity. Entropy measures the complexity or uncertainty of a signal. For instance, the entropy of signal X is defined by

$$H(X) = -\sum_{x \in X} p(x) log_2(p(x)) \tag{4.2}$$

If there is another signal Y, then the joint entropy of signal X and Y is defined as

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) log_2(p(x,y)) \tag{4.3}$$

The mutual information measures the dependence between the two signals. The mutual information between X and Y is defined as

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{4.4}$$

Also, the mutual information depends on the number of observations in signals. To remove this influence, the symmetrical uncertainty is utilized to normalize the mutual information from 0 to 1, which is illustrated below.

$$symmetrical\ uncertainty = \frac{2 \times I(X;Y)}{H(X) + H(Y)} \tag{4.5}$$

Besides, the selection criterion of this method is based on an evaluation function. The evaluation function is a combined test which consists of the correlation between features and class and the inter-correlation between features. It can be described as follows.

$$M_S = \frac{n \cdot r_{cf}}{\sqrt{n + n(n-1)\overline{r_{ff}}}} \tag{4.6}$$

where n is the size of the features; $\overline{r_{ff}}$ is the mean value of the feature to feature correlations; $\overline{r_{cf}}$ is the mean value of the feature to class correlations and M_S is the ranking criterion of a feature subset. From equation 4.6, we can be see that this evaluation considers both the feature to class correlation and the inter-correlation between features. The value of evaluation function gives a ranking on the feature subsets. After ranking by the evaluation function, irrelevant features can be removed due to their low correlations with the class and redundant features can also be eliminated because of their high correlations with the remaining features. The diagram of this CFS algorithm is shown in Figure 4.2.

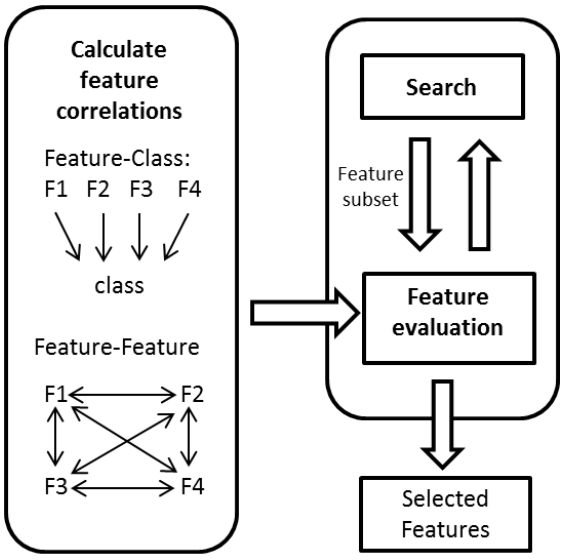


Figure 4.2: The diagram of CFS algorithm

Forward selection was used as the search strategy in this method. The implementation of this method is described as follows. The search method starts with the empty

set without features inside. Each additional feature added to the empty set is calculated by the evaluation function and the feature with highest score is ultimately added to the subset. The second step includes testing each of the remaining features added to the subset and choosing the best one with the highest score. Similarly, the next step is the repetition of the second step which tries to find next best feature from the remaining features. The searching procedure will continue until the last remaining feature was added to the subset. After that, the subset with the maximum score of the evaluation function is selected [27]. Table 4.1 gives an example of CFS applied to a dataset with four features, which illustrates a forward feature selection search through the whole feature subset based on the evaluation function.

Feature subset	k	M_s
[ ]	0	0
[Feature 1]	1	0.12
[**Feature 2**]	1	**0.18**
[Feature 3]	1	0.04
[Feature2 Feature 1]	2	0.15
[**Feature2 Feature 3**]	2	**0.22**
[Feature2 Feature 4]	2	0.13
[Feature2 Feature3 Feature1]	3	0.16
[**Feature2 Feature3 Feature4**]	3	**0.28**
[Feature2 Feature3 Feature4 Feature1]	4	0.26

Table 4.1: Forward feature selection using the result of evaluation function

k is the size of the feature subset, M_s is the evaluation function in equation 4.6. The subsets in bold have the best performance in each iteration with respect to M_s. The best feature subset Feature 2, Feature 3 and Feature 4 is ultimately returned with the highest value of M_s. After that, the dimension of features is reduced to contain only the features selected by CFS algorithm and the reduced features can then pass a machine learning algorithm for classification and prediction.

4.2.3 Recursive feature elimination with Support Vector Machines

Support vector machine (SVM) is a popular classification algorithm against other algorithms related to computational simplicity. The SVM classifier aims to compute the maximum margin between classes. The decision boundary (hyperplane) is positioned to leave the largest margin between classes. In general, the weights of the decision function are calculated based on the observations lie on the margin, which are defined as support vectors. The support vectors can be utilized to determine the final decision boundary.

The SVM based recursive feature elimination (SVM-RFE) is a backward feature reduction method. The weights w in SVM represent the contributions of each feature to the classifier. In each iteration, the selected feature subset was utilized to train the classifier and compute the weight vector of the features. The feature with smallest weight is eliminated because of its small influence on the weight vector norm. [28]

The whole algorithm is shown below. In the SVM-RFE algorithm, the classification

Algorithm 1 SVM-RFE

Input:
Training dataset: $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}]^{\mathrm{T}}$
Labels of class: $\mathbf{y} = [y_1, y_2, ..., y_n]^{\mathrm{T}}$
Initialization:
Subset of selected features: $\mathbf{s} = [1, 2, ..., \mathrm{i}]$
Output: The performance with each set of features.

1: Update the training dataset with the selected features: $\mathbf{X_s} = \mathbf{X}(:, \mathbf{s})$.
2: Training phase: $\alpha = \mathrm{SVM}(\mathbf{X_s}, \mathbf{y})$.
3: Calculate the weights in the classifier: $\mathbf{w} = \sum_n \alpha_n y_n \mathbf{x_n}$.
4: Evaluate the performance of the classifier for dataset $\mathbf{X_s}$.
5: Compute the ranking criteria: $r_i = (w_i)^2$.
6: Find the feature with smallest r_i: $f = \mathrm{argmin}(\mathbf{r})$
7: Remove the found feature: $\mathbf{s} = \mathbf{s}(1, 2, ..., f-1, f+1, ..., \mathrm{length}(\mathbf{s}))$.
8: Rerun the previous procedure until the feature subset s is empty.

model needs to be trained on each feature subset and its performance computed in each iteration. The output is the performance with each set of features generated from recursive elimination of the least representative feature. After this algorithm, the feature subset with the best performance will be selected.

4.3 Feature selection result

The feature selection methodology is introduced in the previous sections. Filter and wrapper methods are combined in our work. In this section, the feature selection results are described and illustrated.

Firstly, the filter method we used is based on the combination of Wilcoxon rank sum test and CFS algorithm. There is an obvious shortcoming for the CFS algorithm. If the irrelevant feature has low inter-correlation with other features, it can still be selected by CFS algorithm. In case of that, we use the Wilcoxon test before the CFS algorithm. The diagram of the whole process is shown in Figure 4.3.

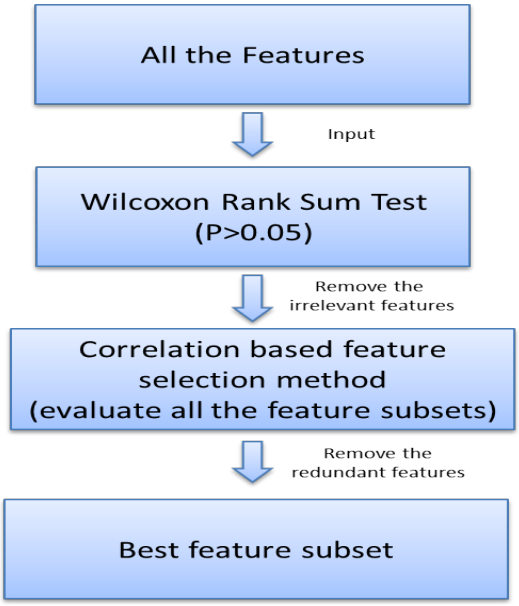


Figure 4.3: The diagram of the filter feature selection algorithm

In this work, 54 features were selected after the Wilcoxon test because their corresponding p-value is less than 0.05. Only 4 features were removed. After this filter feature selection method, ten features were ultimately selected in view of its largest value in the evaluation function 4.4. The ranking criterion M_S in the evaluation function of each feature subset is shown in Figure 4.4.

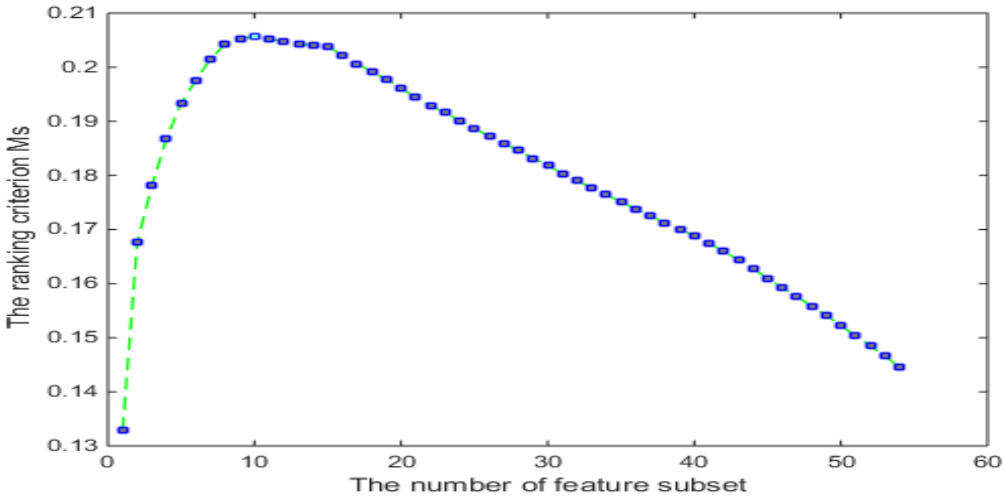


Figure 4.4: The ranking criterion M_s of each feature subset

The green point of Figure 4.4 is the selected feature subset with the number of ten because of its highest M_s value. The ten selected features are based on four feature extraction algorithms which are shown in Table 4.2 : 1) Features of raw ECG signals; 2) Principal component analysis; 3) Autocorrelation function; 4) Discrete wavelet

transform.

Algorithm	Selected Features
Features of raw ECG signals	median absolute deviation
Principal component analysis	second eigenvalue ratio
Autocorrelation function	first zero point location
Discrete wavelet transform	ApEn: level 1 2 3 and 6 detail coefficients Mean: level 6 approximation coefficients Standard deviation: level 2 and 3 detail coefficients

Table 4.2: Feature selection result of filter method

We can see that most of the selected features come from Discrete wavelet transform algorithm. The reason is that the features of DWT algorithm capture both the information from time and frequency domain. Simultaneously, the computational complexity of DWT algorithm is much higher than the other algorithms. In other words, the features based on DWT algorithm performs well at a cost of computational time.

For the wrapper feature selection method, the SVM based RFE algorithm was also used in this work. In the procedure of this algorithm, the performance evaluation method was the average accuracy of 5-fold cross validation in each iteration. The size of elimination in SVM-RFE algorithm corresponds to the number of features to remove at each iteration, which was chosen as one. In other words, only one feature would be eliminated from the whole feature subset during each iteration. The performance of the selected feature subset in each iteration is shown in Figure 4.5.

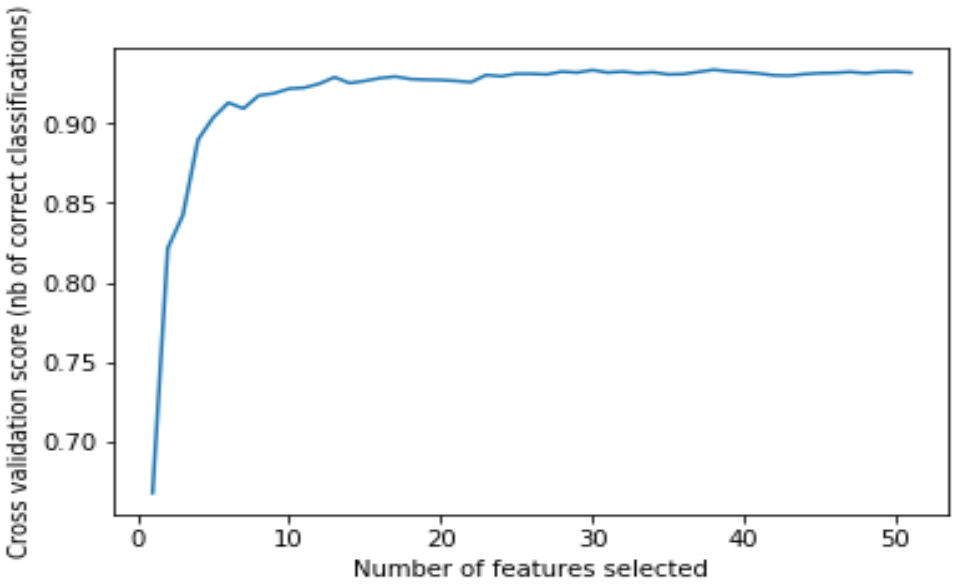


Figure 4.5: The performance of selected feature subset in each iteration

The maximum cross validation accuracy with a value of 0.934 is reached when 38 features were selected in this algorithm. All the selected features are shown in Table

4.3.

Algorithm	Selected Features
Features of raw ECG signals	mean, standard deviation, median absolute deviation, ApEn, kurtosis, skewness
Features of FFT on ECG signals	mean, max, standard deviation, skewness
Principal component analysis	second, third and fifth eigenvalue ratios
Autocorrelation function	first zero point location, zero-crossing rate, first minimum location and amplitude
Discrete wavelet transform	ApEn: level 1 to 5 detail coefficients and level 6 approximation coefficient Mean: level 2 to 6 detail coefficients Standard deviation: level 1 to 4 and 6 detail coefficients, level 6 approximation coefficient Median absolue deviation: level 1, 4 and 6 detail coefficients

Table 4.3: Feature selection result of wrapper method

From the Table 4.3, we can see that much more features were selected in this algorithm compared with the filter method. In the filter method, the features are evaluated by the attributes of feature itself like consistency and statistical distribution. The features which have a high correlation with others are also removed from the subset because of the redundancy. Thus, only a small number of features will be ultimately selected. However, the evaluation criterion is the performance of features in the classifier for the wrapper method, regardless of the redundancy in the feature subset. The combination of features which achieves the highest accuracy will be ultimately selected. Both the filter and wrapper method have their own advantages and these two methods can be applied for different objectives. The performance of the features selected by these two method for classification model will be illustrated in Chapter 5.

ECG signal quality classification with SVM classifier

5

5.1 Introduction

Chapter 3 and 4 introduced the feature extraction and selection methods from ECG signal. This chapter discusses the classifier for identifying the quality of ECG signals. The supervised learning algorithm is an approach where the classification model learns from the datasets and the given labels and then uses this trained model to identify new observations. Different types of classifiers like Naive Bayes, Random Forest, Decision trees and Nearest Neighbours can be used in the supervised learning problems. Support vector machine algorithm was used in this work with its advantage like high accuracy, nice theoretical guarantees regarding overfitting and the acquisition of prediction probability.

5.2 Theoretical Background

Support vector machine (SVM) is widely applied in machine learning algorithm that develops a hyperplane as the decision boundary to separate different class of observations, which is widely used in data classification, image recognition and text categorization [29]. In general, the long distance from the observations of different class to the boundary indicates a good separation between them in SVM. For instance, a dataset of N points x_i, y_i are given, where y_i are the labels indicating the class to each point x_i. The role of SVM is to construct the two parallel hyperplanes that separate different groups of data and make the distance from the hyperplanes to the nearest observations of each group is maximized. The hyperplane is defined as

$$f(x) = w^T x + b = 0 \tag{5.1}$$

where w is the weight vector for the hyperplane and b is the bias. For the binary classification in SVM, if the data can be separated linearly, two parallel hyperplanes called hard margin are generated to separate two groups of observations. The objective is making the distance between the hard margin as large as possible. The maximum-margin hyperplane is defined as the hyperplane which lies in the middle of hard margin. The hard margin is generated from the data points lie on it, which are called support vectors. These two parallel hyperplanes are illustrated as $\vec{w}\cdot\vec{x}+b = 1$ and $\vec{w}\cdot\vec{x}+b = -1$ respectively. The observations that below or above the margin will be classified into two classes. The distance inside the margin is computed as $\frac{2}{\|\vec{w}\|}$. Thus, maximizing the distance is equivalent to minimize $\|\vec{w}\|$. Besides, the constraint should be added to avoid data points that are falling into the margin as

$$y_i(\vec{w}\cdot\vec{x_i} + b) \geq 1,\ 1 \leq i \leq n \tag{5.2}$$

The above hard-margin SVM problem is shown in Figure 5.1.

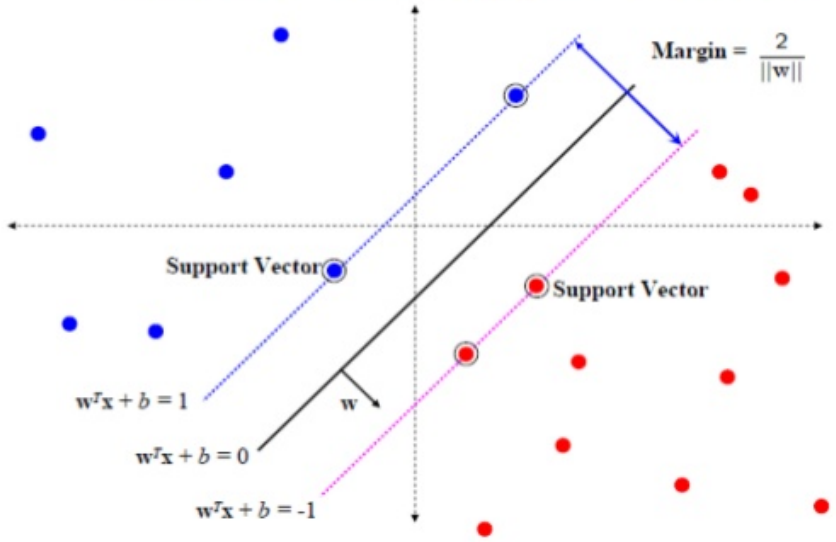


Figure 5.1: Hard margin support vector machine [2]

Thus, the optimization problem can be formulated as minimizing the equation below.

$$J(w) = \frac{w^T w}{2} = \frac{\|w\|^2}{2} \tag{5.3}$$

subject to

$$y_i(w^T x_i + b) \geq 1, \ 1 \leq i \leq n \tag{5.4}$$

To solve the optimization problem with constraint, the Lagrange function is constructed in equation 5.5.

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1] \tag{5.5}$$

where the α_i are called the Lagrange multipliers. The Lagrangian function can be solved by setting the gradient of it to 0. By transforming the above function to Lagrangian dual problem, the simplified optimization problem is obtained as

$$\max \ f(\alpha_1 ... \alpha_n) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i (y_i y_j x_i^T x_j) \alpha_j \tag{5.6}$$

subject to

$$\sum_{j=1}^{n} \alpha_i y_i = 0, \ \alpha_i \geq 0 \tag{5.7}$$

If the data can not be separated linearly, the constraint is violated by introducing a new non-negative variable ξ_i as follows.

$$y_i(w^T x_i + b) \geq 1 - \xi_i \tag{5.8}$$

$$\xi_i \geq 0, \ \ i = 1, ..., N \tag{5.9}$$

For $\xi_i \geq 0$, the observations will fall into the margin. The concrete location of these observations depends on the value of ξ_i. For $\xi_i \leq 1$, the observations will fall on the correct side of the hyperplane. In other case, the observations will fall on the incorrect side of the hyperplane. The optimization problem with variable ξ_i is called soft-margin SVM. Figure 5.2 illustrates the distribution of the observations based on ξ_i.

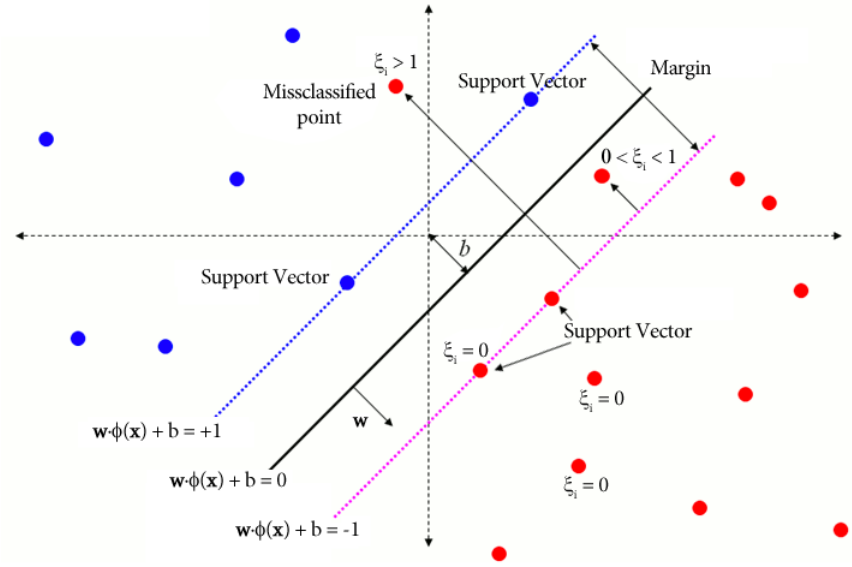


Figure 5.2: Soft margin support vector machine [2]

In soft-margin SVM, the hinge loss function is added in the cost function which are shown below.

$$J(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \tag{5.10}$$

The constraint conditions are same as equation 5.8 and 5.9. C is the parameter which regulates the balance between the number of misclassified observations and the complexity of the machine. The soft-margin SVM could choose decision boundary that has some training error and thus is less likely to overfit the model. An example of how C parameter controls the decision boundary is shown in Figure 5.3.

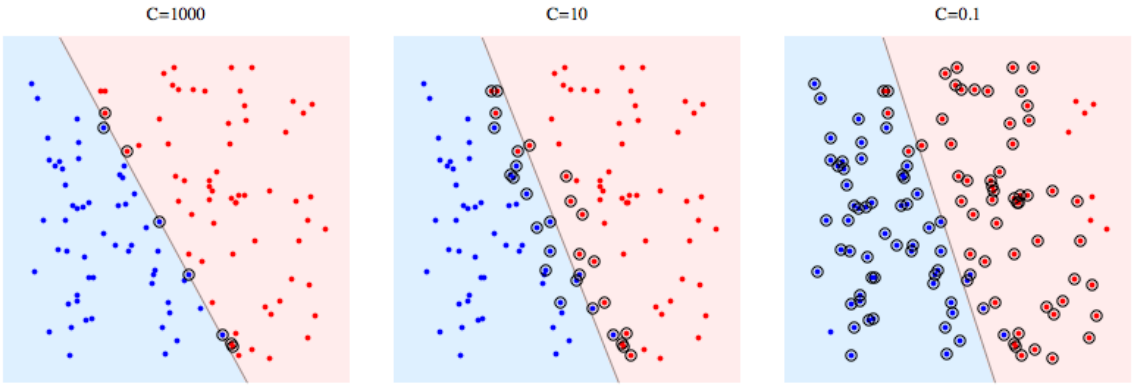


Figure 5.3: The influence of C parameter for the decision boundary [3]

The circled points are support vectors. We can see that decreasing C parameter causes classifier to sacrifice linear separable accuracy in order to gain stability. Thus, tuning C parameter is a crucial procedure in SVM classification problem. Similarly, the dual problem can be formulated by using the Lagrange multipliers in this case as

$$\max f(\alpha_1 ... \alpha_n) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i (y_i y_j x_i^T x_j) \alpha_j \tag{5.11}$$

subject to

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C \tag{5.12}$$

If the original data can not be separated linearly, the projecting function φ is required to project the data to higher dimensions. This allows the algorithm to fit a hyperplane in the projected data space. Then the optimization problem becomes

$$\min J(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \tag{5.13}$$

subject to

$$y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i \tag{5.14}$$

$$\xi_i \geq 0, \ i = 1, ..., N \tag{5.15}$$

The dual problem is also altered by replacing x by $\varphi(x)$ in equation 5.11.

$$\max f(\alpha_1 ... \alpha_n) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i (y_i y_j \varphi^T(x_i) \varphi(x_j)) \alpha_j \tag{5.16}$$

subject to the same constraints

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C \tag{5.17}$$

The kernel function is related to the projecting function φ and defined as follows.

$$k(x_i, x_j) = \varphi^T(x_i)\varphi(x_j) \tag{5.18}$$

Finally the dual problem becomes

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{5.19}$$

subject to

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C \tag{5.20}$$

The weight vector w of the hyperplane with a kernel function is computed as $w = \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i)$. The coefficients α_i can be calculated by solving the quadratic programming problem. Finally, the decision function is generated to identify the new input data z as

$$f(z) = sgn(w^T \varphi(z) + b) = sgn(\sum_{i=1}^{n} y_i \alpha_i k(z, x_i) + b) \tag{5.21}$$

5.3 Performance Evaluation

In this work, ten subjects including fifteen hours data are used to classify the quality of ECG signals. The confusion matrix is a widely used method for performance evaluation in supervised learning problem. In predictive analysis, the confusion matrix consists of four parameters which represent the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) respectively [30]. TP measures the ratio of actual positives that are corrected predicted; TN measures the proportion of actual negatives that are corrected predicted; FP measures the ratio of actual negatives which are incorrectly predicted; FN measures the proportion of actual positives which are wrongly identified as negatives. For binary classification problem, the confusion matrix is shown in Table 5.1.

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	True Positives	False Negatives
Class 2 Actual	False Positives	True Negatives

Table 5.1: The confusion matrix for binary classification

It allows more detailed analysis in performance evaluation than merely the accuracy which is the proportion of correct classifications and can be written as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.22}$$

Sensitivity is the extent to which the actual positives are correctly classified, which is defined by

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5.23}$$

Specificity is the quantification for the actual negatives defined by

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5.24}$$

Accuracy is not enough to represent the performance of the classifier, especially for the unbalanced data sets (the numbers of observations in different classes vary a lot). The accuracy can not reflect the performance of the class with small number of observations, because it is the proportion of correct classifications in this class which depends on the observation number. Therefore, another common used parameters like sensitivity, specificity and AUC from ROC curve are applied in the performance evaluation. The parameter true positive and false positive are two variables in the ROC curve. The ROC curve is plotted based on these two variables at different thresholds, which illustrates the diagnostic ability of a binary classifier across all possible distinction thresholds. AUC (Area under the ROC curve) is a parameter for measuring the performance of classifier when the discrimination threshold is changed [31].

K-fold cross validation is utilized to estimate the performance and stability of the classifier in this work. Cross-validation is primarily applied in supervised learning problem, which can be used to check the overfitting. It will produce a less biased estimation of the model than other methods like a simple train-test split method [32]. The general procedures are shown below:

1. Mix the dataset.

2. Uniformly divide the dataset into k groups.

3. Take one group as test dataset and the remaining groups as training dataset.

4. Fit a model on the training dataset and evaluate its performance on the test dataset.

5. Repeat above steps until all the groups are selected.

The scheme of a 5-fold cross validation procedure is shown in Figure 5.4.

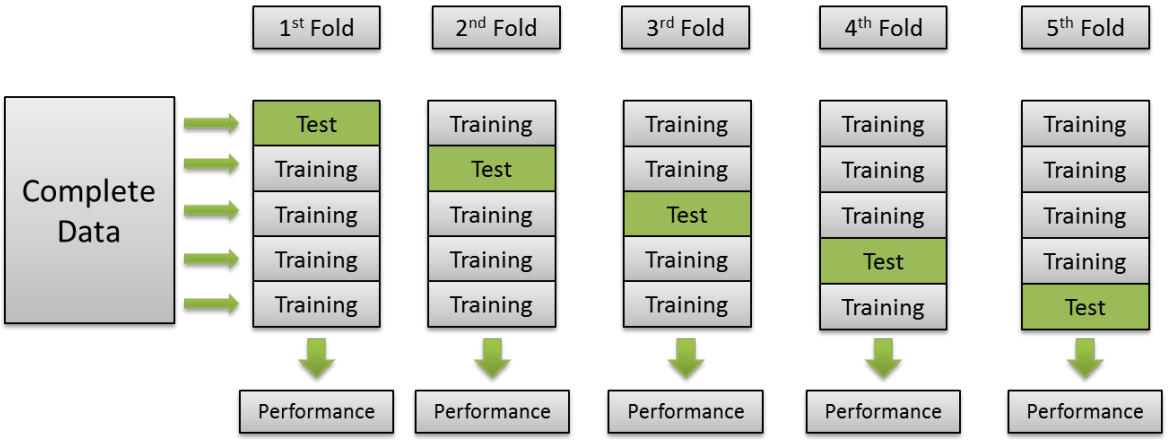


Figure 5.4: The scheme of 5-fold cross validation

In each iteration, the parameters like accuracy, sensitivity and specificity quantifying the performance of classifier on test dataset is calculated. Finally, the average value of these parameters represents the absolute performance and the variance of it represents the stability of the classification model. Thus, cross-validation can also be used to check whether the model has been overfitted. The overfitting problem indicates that the trained model is not sensitive to new dataset. In this case, the classifier will achieve a bad performance for new dataset even if it has a high performance for the training data.

5.4 Experimental results

Before performing the classification experiment, 25 % of datasets are extracted as validation data. The remaining 75 % of datasets are applied into k-fold cross validation to assess the performance and stability of the classification model. The parameter k is chosen as five (5-fold). In addition, tuning the penalty parameter C of SVM classifier is required in this work because of the unbalanced training data (the number of observations in each class is different). For the balanced data set, the penalty parameter C of each class in the optimization function is equal (see in equation 5.13). However, this optimization function is changed for the unbalanced data set as follows.

$$\min\ J(w,\xi) = \frac{1}{2}w^T w + C^+ \sum_{i \in C^+} \xi_i + C^- \sum_{j \in C^-} \xi_j \tag{5.25}$$

where C^+ is the class labeled as 1 and C^- is the class labeled as 0. To tuning the parameter C, its value is adjusted inversely to class frequencies in the input data. In this case, the SVM classifier will give higher penalty to the class with less number of observations, which can improve the classification accuracy in this class. The 5-fold cross validation generates five classification models in total and the best one of them is the model with the maximum average of accuracy, sensitivity and specificity. The best model is then applied in the prediction work of the validation data.

5.4.1 Classification performance for filter feature selection method

As exposed in Chapter 4, two feature selection methods are used in this work. For filter method, ten features were ultimately selected (show in Table 4.2). The features were then utilized to train the SVM classifier with 5-fold cross validation. The evaluation criterion of performance depends on accuracy, sensitivity, specificity, f1-score of confusion matrix and ROC curve of the classification model. The performance based on confusion matrix is shown in Table 5.2.

Performance	Average value	Standard deviation
Accuracy	0.920	0.053
Sensitivity	0.952	0.027
Specificity	0.891	0.102
F1-score	0.919	0.047

Table 5.2: The performance of 5-fold cross validation for features from filter method

It can be seen that the performance of confusion matrix is acceptable as all the values are over 0.89. The performance based on ROC curve in each iteration of cross validation was also plotted in Figure 5.5. The AUC measures the performance of the classification model. The statistical features of AUC were both annotated in the plot.

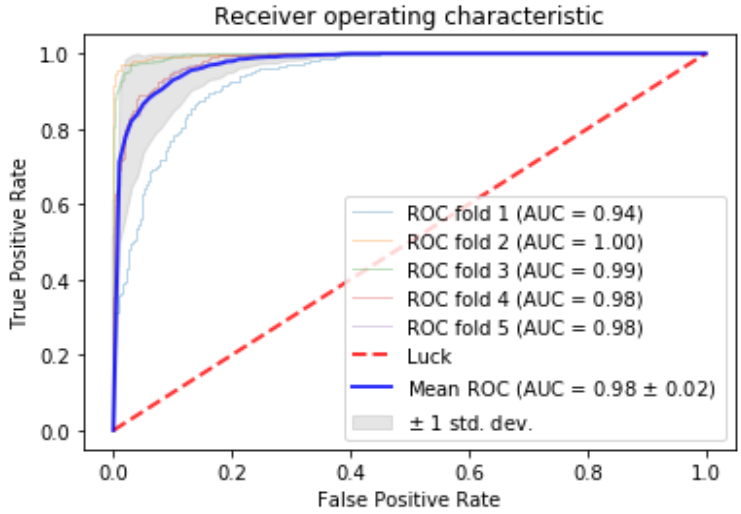


Figure 5.5: The ROC curve of cross validation for filter method

From Figure 5.5 we can see the value of AUC is near one in all 5 iterations of cross validation, which illustrates that the training model performs well in separating the good and bad quality observations. It also does not show over-fitting problem related to the small value of standard deviation. After the cross validation, five models are generated and the second model is chosen as the best model with the highest accuracy and F1-score. Then we use this model to make predictions of the new validation data. The performance of this classification model is shown in Table 5.3.

Performance	Average value
Accuracy	0.916
Sensitivity	0.837
Specificity	0.949
F1-score	0.856
AUC	0.967

Table 5.3: The performance of filter method for validation data

From the performance of the selected model for validation data, we can see that the evaluation parameters are all above 0.8. The sensitivity and F1-score are lower, meaning that the prediction error for bad quality observations is higher compared with the training phase.

5.4.2 Classification performance for wrapper feature selection method

38 features were selected based on wrapper method which are shown in Table 4.3. The classification performance was achieved by using the same training procedure as mentioned before. Table 5.4 shows the performance of 5-fold cross validation with these 38 features.

Performance	Average value	Standard deviation
Accuracy	0.942	0.039
Sensitivity	0.954	0.007
Specificity	0.930	0.062
F1-score	0.940	0.036

Table 5.4: The performance of 5-fold cross validation for features from wrapper method

The ROC curve in each iteration of cross validation was also plotted in Figure 5.6.

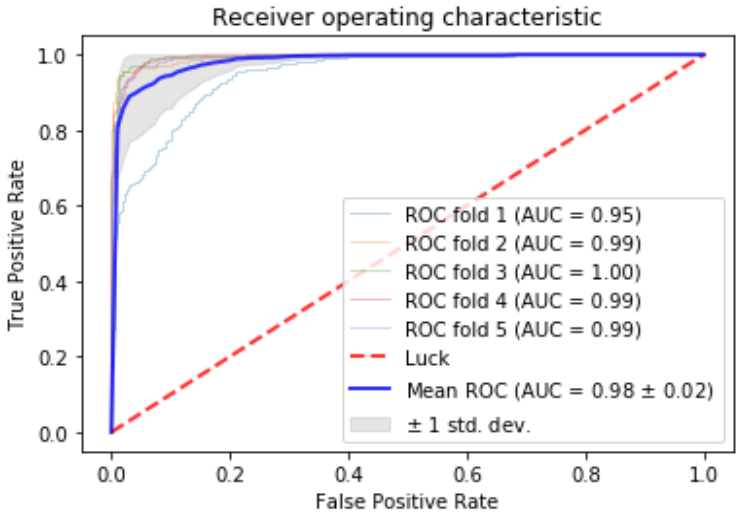


Figure 5.6: The ROC curve of cross validation for wrapper method

From Table 5.4 and Figure 5.6, we can see the mean and standard deviation of AUC for wrapper method are equivalent to the filter method. The performance based on confusion matrix of wrapper method is improved compared with the filter method. Also the performance of wrapper method in training phase is more stable considering the standard deviation. In this case, the second model is chosen as the best model with the highest accuracy and F1-score. Then we use this model to predict labels of the validation data. The performance of this model is shown in Table 5.5.

Performance	Average value
Accuracy	0.924
Sensitivity	0.943
Specificity	0.916
F1-score	0.882
AUC	0.981

Table 5.5: The performance of wrapper method for validation data

Compared the performance in Table 5.3 and 5.5, we find that almost all the evaluation parameters are improved by the wrapper feature selection method. The model is accurate enough to classify the quality of ECG signals on new subjects.

5.4.3 Classification performance for hybrid feature selection method

If we combine the filter and wrapper feature selection method as hybrid method, only four features were selected which are shown in Table 5.6.

Algorithm	Selected Features
Features of raw ECG signals	median absolute deviation
Discrete wavelet transform	ApEn: level 3 detail coefficients Standard deviation: level 2 and 3 detail coefficients

Table 5.6: Feature selection result of hybrid method

Table 5.7 shows the performance of 5-fold cross validation with these four features.

Performance	Average value	Standard deviation
Accuracy	0.923	0.044
Sensitivity	0.946	0.039
Specificity	0.902	0.091
F1-score	0.922	0.040

Table 5.7: The performance of 5-fold cross validation for features from hybrid method

The corresponding ROC curve of each iteration in cross validation was plotted in Figure 5.7.

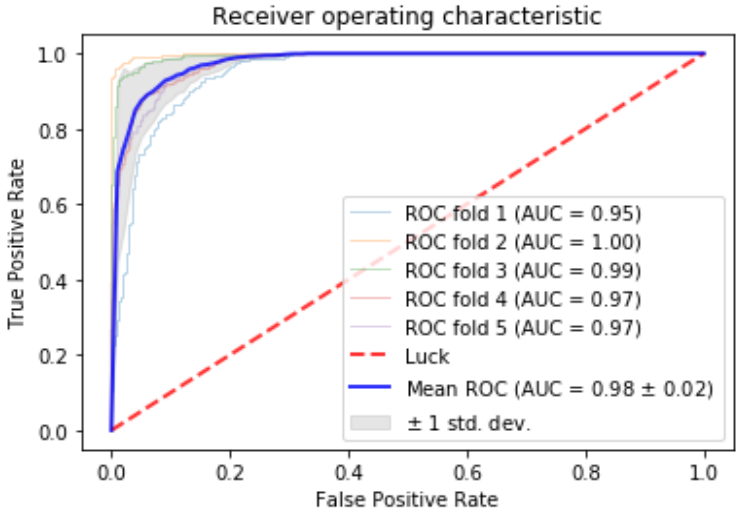


Figure 5.7: The ROC curve of cross validation for hybrid method

Also, the performance of best training model on validation data is shown in Table 5.8.

Performance	Average value
Accuracy	0.918
Sensitivity	0.876
Specificity	0.936
F1-score	0.865
AUC	0.974

Table 5.8: The performance of hybrid method for validation data

From Figure 5.7, Table 5.7 and 5.8, we can see the mean and standard deviation of AUC for hybrid method are equivalent to the other methods. The performance based on the confusion matrix of the hybrid method is improved compared with the filter method but not better than the wrapper method.

In conclusion, all of these three feature selection methods achieve good performance. The choice of the method appears the trade-off between accuracy and computational complexity. In this work, we choose the hybrid method because only four features have remained. It tremendously reduces the computational time in the real application but still performs well.

5.4.4 Classification performance for each feature extraction method

To compare the proposed feature extraction methods, we also compute the classification performance of each set of features. Table 5.9 shows the mean and standard deviation of performance on different feature extraction methods in the cross validation phase.

Features	Acc	Sen	Spe	F1	AUC
Raw	0.861+0.061	0.905+0.047	0.822+0.101	0.862+0.055	0.92+0.07
ACF	0.741+0.023	0.843+0.083	0.651+0.059	0.752+0.031	0.83+0.05
PCA	0.819+0.030	0.899+0.089	0.749+0.062	0.822+0.038	0.92+0.03
DWT (ApEn)	0.899+0.070	0.907+0.074	0.892+0.072	0.895+0.072	0.96+0.04
DWT (Mean)	0.658+0.039	0.728+0.257	0.595+0.170	0.644+0.110	0.78+0.09
DWT (Std)	0.804+0.098	0.840+0.172	0.773+0.047	0.793+0.125	0.89+0.07
DWT (MAD)	0.805+0.089	0.821+0.103	0.792+0.177	0.801+0.078	0.90+0.09

Table 5.9: Performance comparison of the proposed feature extraction methods in cross validation phase

The performance comparison between the proposed feature extraction methods for the validation data is shown in Table 5.10.

Features	Acc	Sen	Spe	F1	AUC
Raw	0.756	0.872	0.610	0.692	0.865
ACF	0.711	0.760	0.663	0.635	0.781
PCA	0.805	0.868	0.725	0.738	0.875
DWT (ApEn)	0.925	0.892	0.941	0.876	0.948
DWT (Mean)	0.710	0.858	0.612	0.645	0.853
DWT (Std)	0.816	0.861	0.798	0.742	0.898
DWT (MAD)	0.782	0.871	0.714	0.726	0.872

Table 5.10: Performance comparison of the proposed feature extraction methods for validation data

From Table 5.9 and 5.10, we can see that the ApEn of DWT achieves the best performance because of its high computational complexity. The ApEn needs to compare the similarity between succeeding patterns of ECG signals. The calculation of ApEn considers each fragment of the given signals. The statistical features standard deviation and median absolute deviation are not better than ApEn but still achieve a satisfying performance in both training and validation phase. The features of the PCA algorithm and raw ECG signals both perform well, but the performance of the PCA algorithm is more stable than the features from raw signals. The ACF method does not achieve a good performance compared with other methods. The reason of that is the features of ACF are sensitive to the low-frequency noise. In other words, even small low-frequency variations of ECG amplitude can change the shape of ACF. Thus, some good quality ECG signals with small low-frequency variations can be mistakenly classified in this case.

In conclusion, the features of the DWT algorithm achieve the best performance because of its high computational complexity. It considers both the temporal and spectral information of ECG signals. The PCA algorithm also performs well with only five features and low computational complexity. The performance of features on raw ECG signals is acceptable but not stable enough. The ACF method still requires to be improved in the future.

5.4.5 Discussion

The classification results illustrate that the performance of features selected by the filter, wrapper and hybrid methods are both over 0.8. On the other hand, the performance of wrapper method is higher compared to the other methods, especially for the validation data. But the filter method has the advantage of low computational time and low redundancy. The features are selected based on the feature to class and feature to feature correlations regardless of the classifier. After that, the irrelevant and redundant features are eliminated, which also reduces the dimensions of the whole dataset. For the wrapper method, it is a classifier based feature selection method. The performance is generated by training the classifier for each feature subset and the features are ultimately selected with the best performance. Because of that, the total computational time is increased, but the selected features achieve the best performance for a certain classifier. As we mentioned in Chapter 4, 38 features are selected by the wrapper method in contrast to 10 features by the filter method. It indicates that more information of the ECG signals has been captured by these 38 features and the classifier can distinguish the quality of ECG signals on different dimensions. This is why the performance of wrapper method is better than the filter method. For the hybrid method, it combines the advantage of both filter and wrapper methods. The performance has been improved on average compared with the filter method but with fewer features. In other words, the hybrid method can achieve a satisfying performance with low computational time in the training phase.

In general, all of these three feature selection methods can be used in the classification problem. Each method has its own advantages. In real application, our objective is designing a ECG signal quality indicator with low computational time and acceptable performance. Thus, we choose the hybrid method as the feature selection method.

5.5 ECG signal quality prediction result

The SVM classifier will produce a posterior probability of each observation after the training phase, which is very useful in practical predictions. The posterior probability gives a degree of certainty about the classification result. In this work, the ECG quality can be predicted based on the posterior probability. As we mentioned before, the output of the SVM classifier is

$$f(x) = h(x) + b \tag{5.26}$$

where

$$h(x) = \sum_i y_i \alpha_i k(x_i, x) \tag{5.27}$$

To map the output of SVM to probabilities, the Platt scaling algorithm is used to solve this problem. It produces the posterior probability estimates based on a sigmoid function as shown below [33].

$$P(y = 1|x) = \frac{1}{1 + exp(Af(x) + B)} \tag{5.28}$$

where A and B are two parameters which are generated by this algorithm, f(x) is the output of SVM classifier. The range of the posterior probability is from 0 to 1, which indicates the predictions can now be made according to the rules that y = 1 if $P > 0.5$ and y = 0 if $P < 0.5$. For linear SVM classification, a flat hyperplane is generated as the decision boundary to distinguish different classes. In this case, the posterior probability of SVM classifier represents the distance to the decision boundary. The high value of probability for good quality signals indicates that these observations are far from the decision boundary.

In this work, the ECG signals are predicted as three quality levels depending on the posterior probability computed by the SVM classifier. The prediction rules are shown as follow.

1. Perfect quality: the posterior probability is over 0.5.
2. Average quality: the posterior probability is between 0.001 and 0.5.
3. Bad quality: the posterior probability is below 0.001.

We use the best three features based on Fisher score method for visualization of the prediction result of SVM with linear kernel function. The 3D plot of the observations with its corresponding quality is shown in Figure 5.8.

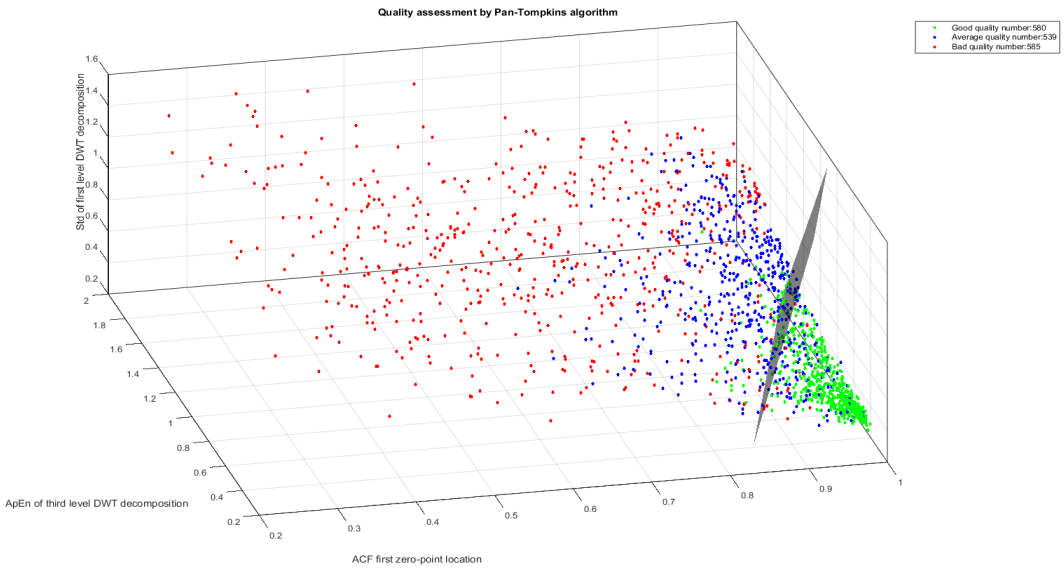


Figure 5.8: The 3D plot of ECG quality prediction result

The annotation rules are that the ECG signal will be labeled as class 1 if it is perfect. Thus, if the probability is over 0.5, the quality will be predicted as perfect. Considering that the observations with poor quality will decrease the probabilities a lot, we make a low threshold 0.001 of probability to separate the average and bad quality ECG signals. Because the ten-second window was used to divide the whole ECG signals, the quality prediction result will display on each window, which is shown in Figure 5.9.

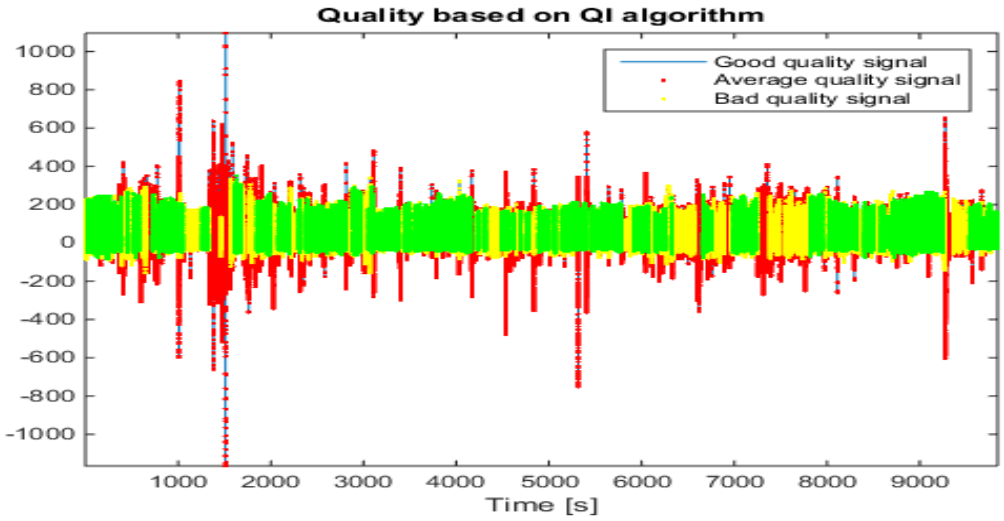


Figure 5.9: The quality prediction result of ECG signals

The green, yellow and red color represent the perfect, average and bad quality separately. The total period of the ECG signals is almost three hours. To make the quality prediction result more clear, we zoom in the figure and plot four segments of it as shown in Figure 5.10.

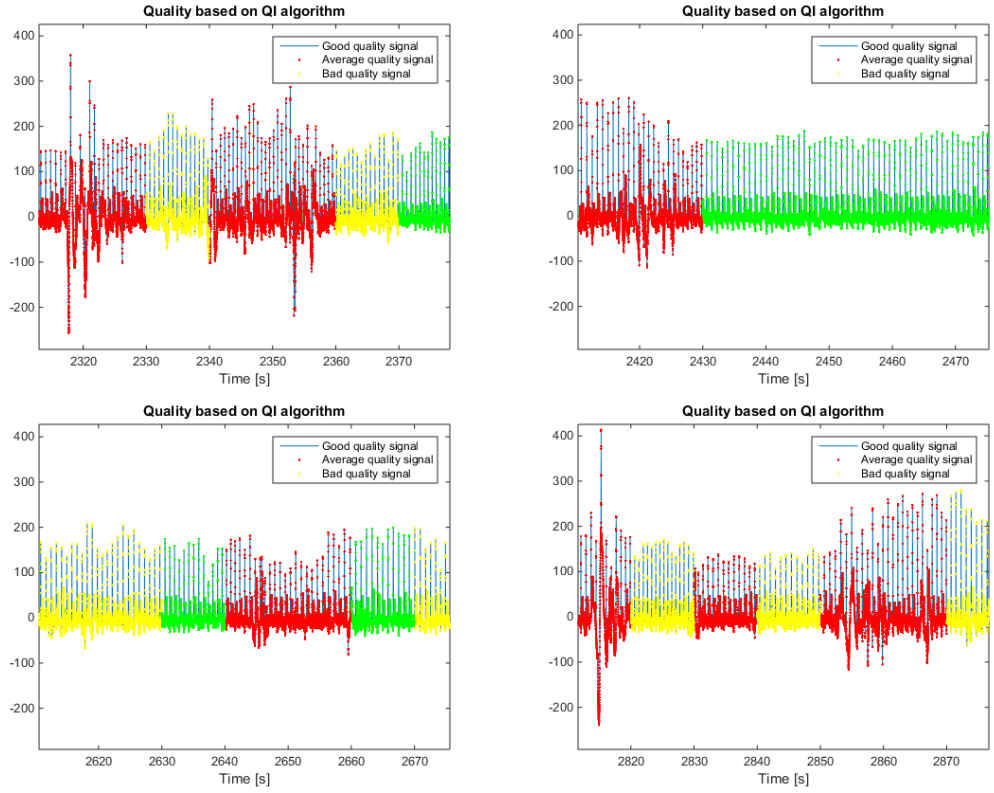


Figure 5.10: The quality prediction result of ECG signals

Apart from that, we also choose another four segments and plot their corresponding prediction result in Figure 5.11.

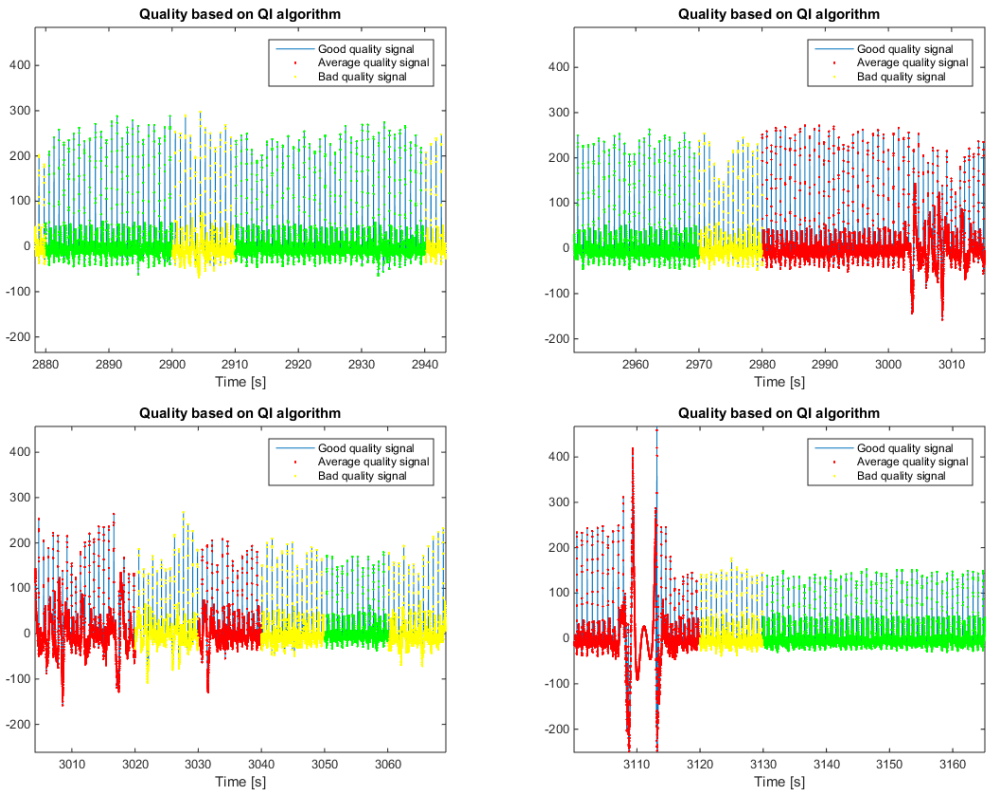


Figure 5.11: The quality prediction result of ECG signals

From Figure 5.10 and 5.11, we can see that the classifier predicts the quality of ECG signals on each ten-second window. The red color parts are ECG signals with bad quality, which includes a lot of variations on the amplitude. The green color parts are perfect ECG signals. The yellow color parts represent ECG signals with intermediate quality, which are not perfect but have only small fluctuations on the amplitude. The perfect ECG signals can be used directly for any ECG analysis work. The intermediate quality ECG signals can be implemented for the applications like heart rate calculation where the morphology of the low frequency components is not relevant. For the bad quality signals, further processing needs to be applied to enhance the quality of these signals.

6 Conclusion and Future Work

6.1 Conclusion

In this work, we aim to design a automatic ECG signal quality indicator with a supervised learning algorithm for wearable devices. The ECG signal is the non-stationary and time series signal which includes a large amount of morphological structures like QRS complex, P wave and T wave. In the practical application, the ECG signals are easy to be contaminated by the noise and artifacts because of the respiration, body movement and muscle activity in the collection. Sometimes these artifacts have overlapping spectra with the ECG waves, which results in the difficulty of identifying these contaminated ECG signals. As we mentioned before, the Discrete wavelet transform (DWT) is a convenient technique for decomposing the non-stationary signals into sub-bands. The temporal and spectral information of the ECG signal can be achieved after this algorithm. Also, the computational complexity of DWT is not high compared with other popular signal decomposition techniques like empirical mode decomposition (EMD) and independent component analysis (ICA). Besides, there are many types of wavelets to choose for different application. In this work, we choose the Daubechies wavelet of order six from the wavelet bank as the utilized wavelet due to its similar morphological structure with the ECG waves. In the reference [34], we can see that the Daubechies wavelet of order six achieves the maximum cross correlation coefficient with the ECG signal, which indicates the ECG signal in each sub-band will remain its original morphological structure as much as possible. The sub-bands can help to identify different ECG waves and noises located at their corresponding frequencies. Because of these advantages of Wavelet decomposition, we develop the first feature extraction method based on this approach. Our second feature extraction approach is based on Autocorrelation function (ACF). The ACF is a technique to measure the similarity between different components of a time series signal. The analysis of autocorrelation is a mathematical tool for finding the periodic information of a signal. The ACF is well suitable for analyzing the ECG signals because of the repeated patterns in this signal. Our last method for extracting features is based on Principal component analysis. In this method, we construct a matrix including each QRS complex detected by the R peaks detector in each column. Then we compute the first three eigenvalue ratios over the sum of all eigenvalues as features by the PCA algorithm. The eigenvalue ratios represent the similarity between the successive QRS complex of ECG signals. The advantage of this method is that it does not demand an accurate R peaks detector. If there are R peaks are wrongly detected, it illustrates that there are noise or artifacts in the ECG signals which can be reflected in the features. Besides, the extracted features are sensitive to the variations in any QRS complex, even a small change of it. Afterwards, it is necessary to select the best features before training a classifier. We use two

feature selection methods named filter and wrapper method to choose the best features. Then the selected features are utilized to train a SVM classifier. The SVM classifier was used due to its high accuracy and low computational complexity. Besides, the SVM classifier can produce a posterior probability of each observation, which is useful in the quality prediction. For example, we can predict the quality of ECG signals into several degrees by thresholding the probabilities. In addition, both of these two feature selection methods achieve high performance, but the performance of wrapper method is better than the other feature selection method at a price of computational complexity.

At last, we compare the classification result of the features from these three algorithms. The DWT algorithm achieves the best performance due to its ability to capture both the temporal and spectral information of the ECG signal. Also, more features are extracted of this algorithm in contrast with other algorithms. The performance of the PCA algorithm is better than the ACF algorithm, and the reason for that is its ability to distinguish small changes in the QRS complex.

6.2 Future work

The designed automated ECG signal quality indicator achieves a decent performance to distinguish the good and bad signals. However, there are still some limitations for this algorithm. At first, the labelling work is focus on the perfect and not perfect ECG signals. After training the classification model, a large amount of ECG segments will be predicted as bad quality, even we set a low threshold of the probability. In real applications like heart rate monitoring, lots of ECG signals need to be eliminated and recovered if they are predicted as bad quality. In this case, we will lose too much information of the signal for analysis. In addition, the other applications for ECG signals like cardiovascular disease diagnosis and arrhythmia recognition are also crucial in our daily life. However, this automated quality indicator is focus on the ECG signals from the healthy people. For the unhealthy people, the quality of these ECG signals might be incorrectly predicted because of its irregular morphological structure of the ECG signals. In the future, this algorithm still needs to be tested for the ECG signals from the unhealthy people. The features based on DWT are the combination of statistical and non-linear features. The non-linear features achieve high performance but at the expense of high computational complexity. The future work can be the proposition of new features with low complexity on the DWT algorithm.

Besides, deep learning algorithm based on neural network can also be used in the quality evaluation. It does not need any feature extraction methods to represent the signal. The input of the training model can be the original temporal or spectral ECG signal including all the information. Thus, the neural network can achieve a high performance but with expensive computation. Except for the ECG signal, the PPG signal is also a famous biomedical signal of the wearable sensors. Both of these two biomedical signals are non-stationary and periodic signals. Our proposed feature extraction methods are based on these properties of the ECG signal. At the same time, these methods might also be suitable for the PPG signals, which demands to be confirmed.

Appendix

A

Definition 1: Continuous wavelet transform of signal $x(t)$ is illustrated below.

$$CWT(i,j) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|a|}} \psi(\frac{t-j}{i}) dt \tag{A.1}$$

where i is the scaling parameter; j is the shifting parameter and ψ is the wavelet function.

Definition 2: Fisher score is a parameter to quantify how well the two samples a and b can be separated.

$$F(a,b) = \frac{(\mu_1 - \mu_2)^2}{s_1^2 + s_2^2} \tag{A.2}$$

where μ_1 and s_1^2 are the mean and variance of class 1; μ_2 and s_2^2 are the mean and variance of class 2.

Features (Detail coefficient)	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Mean	0.206	0.206	0.203	0.205	0.199	0.186
Std	0.070	0.103	0.494	0.547	0.327	0.028
MAD	0.177	0.190	0.238	0.014	0.192	0.056
ApEn	0.265	**0.594**	**0.628**	**0.521**	0.003	0.038

Table A.1: The Fisher score of the DWT features

Features	Fisher score
First local maximum amplitude	**0.130**
First local maximum location	0.067
Second local maximum amplitude	**0.127**
Second local maximum location	0.054
Third local maximum amplitude	0.112
Third local maximum location	0.043
First local minimum amplitude	0.071
First local minimum location	0.029
First zero point location	**0.296**
Zero-crossing rate	0.028
Variance of zero-zero intervals	0.035

Table A.2: The Fisher score of the ACF features

Features (Eigenvalue ratio)	**1st ratio**	**2nd ratio**	**3rd ratio**	4th ratio	5th ratio
Fisher score	**0.585**	**0.559**	**0.428**	0.362	0.319

Table A.3: The Fisher score of the PCA features

Features	Fisher score
Mean	0.207
Std	0.191
MAD	0.179
Kurtosis	0.132
Skewness	0.185
ApEn	0.416

Table A.4: The Fisher score of the features on the original ECG signal

Features	Fisher score
Mean	0.362
Max	0.036
Std	0.286
Kurtosis	0.038
Skewness	0.002
ApEn	0.005

Table A.5: The Fisher score of the FFT features

Bibliography

[1] Marek Gierliski. P-values and statistical tests. Division of Computational Biology, 2018.

[2] Event classification and prediction using svm. https://www.slideshare.net/kambliruta/event-classification-prediction-using-support-vector-machine. Accessed: 2016-04-05.

[3] Influence of c parametr in svm. http://yaroslavvb.com/upload/save/so-libsvm.png. Accessed: 2010-09-30.

[4] Pablo Laguna, Roger G Mark, A Goldberg, and George B Moody. A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg. In *Computers in cardiology 1997*, pages 673–676. IEEE, 1997.

[5] Adam Gacek. An introduction to ecg signal processing and analysis. In *ECG Signal Processing, Classification and Interpretation*, pages 21–46. Springer, 2012.

[6] Christina Orphanidou, Timothy Bonnici, Peter Charlton, David Clifton, David Vallance, and Lionel Tarassenko. Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE journal of biomedical and health informatics*, 19(3):832–838, 2015.

[7] Dieter Hayn, Bernhard Jammerbund, and Günter Schreier. Qrs detection based ecg quality assessment. *Physiological measurement*, 33(9):1449, 2012.

[8] GD Clifford, J Behar, Q Li, and Iead Rezek. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiological measurement*, 33(9):1419, 2012.

[9] Jakub Kužílek, Michal Huptych, Václav Chudáček, Jiří Spilka, and Lenka Lhotská. Data driven approach to ecg signal quality assessment using multistep svm classification. In *Computing in Cardiology, 2011*, pages 453–455. IEEE, 2011.

[10] Jinseok Lee, David D McManus, Sneh Merchant, and Ki H Chon. Automatic motion and noise artifact detection in holter ecg data using empirical mode decomposition and statistical approaches. *IEEE Transactions on Biomedical Engineering*, 59(6):1499–1506, 2012.

[11] Udit Satija, Barathram Ramkumar, and M Sabarimalai Manikandan. An automated ecg signal quality assessment method for unsupervised diagnostic systems. *Biocybernetics and Biomedical Engineering*, 38(1):54–70, 2018.

[12] Hatice Ceylan Koydemir and Aydogan Ozcan. Wearable and Implantable Sensors for Biomedical Applications. In Bohn, PW and Pemberton, JE, editor, *ANNUAL REVIEW OF ANALYTICAL CHEMISTRY, VOL 11*, volume 11 of *Annual Review of Analytical Chemistry*, pages 127–146. 2018.

[13] Jinseok Lee, David D McManus, Sneh Merchant, and Ki H Chon. Automatic motion and noise artifact detection in holter ecg data using empirical mode decomposition and statistical approaches. *IEEE Transactions on Biomedical Engineering*, 59(6):1499–1506, 2012.

[14] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[15] Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

[16] F Buendía-Fuentes, MA Arnau-Vives, Antonio Arnau-Vives, Yolanda Jiménez-Jiménez, J Rueda-Soriano, E Zorio-Grima, A Osa-Sáez, LV Martínez-Dolz, L Almenar-Bonet, and MA Palencia-Pérez. High-bandpass filters in electrocardiography: source of error in the interpretation of the st segment. *ISRN cardiology*, 2012, 2012.

[17] C Saritha, V Sukanya, and Y Narasimha Murthy. Ecg signal analysis using wavelet transforms. *Bulg. J. Phys*, 35(1):68–77, 2008.

[18] Hari Mohan Rai, Anurag Trivedi, and Shailja Shukla. Ecg signal processing for abnormalities detection using multi-resolution wavelet transform and artificial neural network classifier. *Measurement*, 46(9):3238–3246, 2013.

[19] Quanquan Gu, Zhenhui Li, and Jiawei Han. Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*, 2012.

[20] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE Trans. Biomed. Eng*, 32(3):230–236, 1985.

[21] Kanti V Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.

[22] Nazrul Hoque, DK Bhattacharyya, and Jugal K Kalita. Mifs-nd: a mutual information-based feature selection method. *Expert Systems with Applications*, 41(14):6371–6385, 2014.

[23] Alan Jović, Karla Brkić, and Nikola Bogunović. A review of feature selection methods with applications. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, pages 1200–1205. IEEE, 2015.

[24] Sebastián Maldonado, Richard Weber, and Fazel Famili. Feature selection for high-dimensional class-imbalanced data sets using support vector machines. *Information Sciences*, 286:228–246, 2014.

[25] Jean Dickinson Gibbons and Subhabrata Chakraborti. Nonparametric statistical inference. In *International encyclopedia of statistical science*, pages 977–979. Springer, 2011.

[26] R Shier. Statistics: 2.2 the wilcoxon signed rank sum test. *Mathematics Learning Support Centre. Retrieved from http://www. statstutor. ac. uk/resources/uploaded/wilcoxonsignedranktest. pdf*, 2004.

[27] Mark Andrew Hall. Correlation-based feature selection for machine learning. 1999.

[28] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[29] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[30] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

[31] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[32] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[33] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[34] Brij N Singh and Arvind K Tiwari. Optimal selection of wavelet basis function applied to ecg signal denoising. *Digital signal processing*, 16(3):275–287, 2006.