

Atrial Fibrillation classification from a short single lead ECG recording

by

Yuchen Yin

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday September 21, 2020 at 14:00.

Student number: 4797558
Project duration: November 29, 2019 – September 21, 2020
Thesis committee: dr. ir. R. C. Hendriks, TU Delft, supervisor
dr. ir. C. Varon, TU Delft, daily supervisor
dr. ir. F. Fioranelli, TU Delft

Acknowledgments

Reaching this last moment of my thesis project, I feel grateful to the people who accompany me and support me in the last two years. Without them, I can not finish my master's degree.

First, I want to thank my daily supervisor, Dr. Ir. Carolina Varon. She is patient and responsible. During the weekly meeting, she always helps me improve my method and come up with new ideas. With her detailed comments and suggestions, my thesis report is so much better than the first draft.

I also want to thank my supervisor, Dr. Ir. Richard Hendriks, who never gives up on me and helps me improve my interpretation. He reminds me when my progress is slow, and encourages me when I have trouble with my writing.

Last but not the least, I want to express my gratitude to my family and my friends. They are always there for me, sharing my happiness and sorrow. Especially in this difficult year, caring and accompany is much appreciated.

*Yuchen Yin
Delft, September 2020*

Abstract

This thesis focuses on classifying AF and Normal rhythm ECG recordings. AF is a common arrhythmia occurring in millions of people every year, which could lead to blood clots, stroke or even heart failure. When AF is occurring, the P waves are often absent and RR intervals are often irregular.

This thesis proposes a new Poincaré plot based feature that exploits the distribution and position information of the plot. The Poincaré plot can visually analyze the nonlinear aspects of the heart rate dynamics both qualitatively and quantitatively. In this thesis, the Poincaré plot values are first quantized into small bins, which represent whether corresponding states are visited by the system or not, by setting ones or zeros. The bins are then given weights by the masks based on the probability of each state being visited by the system, and the relative position between the bins and the center of the plot. By calculating the element-wise multiplication and summation between the quantized Poincaré plot and the masks, the expected value of the matrix of the quantized Poincaré plot is computed, and the outliers in the plot are emphasized. Therefore, the proposed feature is assumed to have a higher value for the AF rhythms and a lower value for the Normal rhythm.

Instead of RR intervals, the Poincaré plot used in this thesis is also generated from the peak intervals in the autocorrelation function of both ECG and prediction error. The autocorrelation function aims to evaluate the self-similarity of the ECG signals and thus extracts the irregularity of the AF signals.

The dataset used in this thesis comes from the Physionet Challenge 2017, containing 5076 Normal recordings and 758 AF recordings. In total, 21 Poincaré plot based features are used to train the SVM and random forest models, which yields the F1 score of 0.80 and 0.85, respectively. When using features from the same intervals, RR intervals generate the highest F1 score of 0.77 and 0.81, followed by the peak intervals in the autocorrelation of prediction error with the F1 score of 0.74 and 0.78, followed by the peak intervals in the autocorrelation error of ECG with the F1 score of 0.63 and 0.68. Using the minimum redundancy maximum relevance algorithm, eleven features are selected based on their importance. Training the SVM and RF models with these features reaches the F1 score of 0.78 and 0.84, respectively.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Related work	2
1.3	Research objectives	3
1.4	Contribution	4
1.5	Outline	4
2	Data and pre-processing	5
2.1	Data	5
2.2	R peak detection	6
2.3	Pre-processing	7
2.3.1	Lead inversion detection	7
2.3.2	Normalization of the RR intervals	8
2.3.3	Outlier removal	9
3	Feature extraction	11
3.1	Poincaré plot	11
3.2	Commonly used Poincaré plot based features	12
3.2.1	Ellipse-fitting technique.	13
3.2.2	Temporal based features.	14
3.2.3	Count of bins in the Poincaré plot	16
3.3	Weighted count of bins in the Poincaré plot.	17
3.3.1	Mask processing	17
3.3.2	Distribution of bins in the Poincaré plot	18
3.3.3	Position of bins in the Poincaré plot	21
3.3.4	Combination of distribution and position.	22
3.4	Autocorrelation function	23
3.4.1	Autocorrelation of ECG.	23
3.4.2	Autocorrelation of the prediction error	24
4	Feature selection and classification	27
4.1	Minimum redundancy maximum relevance algorithm	27
4.2	Support vector machines.	29
4.2.1	Searching for optimal parameters	31
4.3	Random forest	32
4.3.1	Decision tree	32
4.3.2	Random forest	33
5	Results	35
5.1	Evaluation metrics	35
5.2	Choice of quantization level Q	36
5.3	Comparison of the distribution of the features	36
5.3.1	RR intervals.	37
5.3.2	Peak intervals in autocorrelation of ECG	38
5.3.3	Peak intervals in autocorrelation of prediction error	40
5.4	Classification results	41
5.4.1	RR intervals.	41
5.4.2	Peak intervals in autocorrelation of ECG	42
5.4.3	Peak intervals in autocorrelation of prediction error	43
5.4.4	Combination of RR intervals and peak intervals in autocorrelation function.	44

5.5	Feature selection	45
5.6	M-lagged Poincaré plot analysis.	46
6	Conclusion	49
6.1	Conclusion	49
6.2	Future work	50
	Bibliography	51

Introduction

1.1. Motivation

Atrial fibrillation (AF) is a rapid and irregular heart beat that can lead to blood clots, stroke, heart failure and other heart-related complications. This kind of cardiac arrhythmia occurs in one to two percent of the general population, including six million Europeans, and this number is likely to double in the following 50 years [41].

According to the duration and underlying reasons for the condition, AF can be divided into five different types [2]:

1. Paroxysmal: Heart can return to the normal rhythm within seven days on its own.
2. Persistent: Heart will not return to the normal rhythm on its own and the irregular rhythm can last for longer than seven days.
3. Long-standing: Irregular rhythm lasts for longer than twelve months.
4. Permanent: The condition lasts for indefinite time.
5. Nonvalvular: AF that is not caused by a heart valve issue.

Normally, the heart contracts and relaxes at a regular beat. This starts from the upper chambers of the heart (the atria), where the sinoatrial (SA) node stimulates the electrical wave between 60-100 times per minute to the atrioventricular (AV) node. The electrical pulses are then delayed at the AV node before going to the ventricles, during which period the ventricles finish filling with blood. After that, the ventricles contract and pump blood to the lungs and the whole body [1].

However, during an AF period, the SA node starts the contraction randomly and disorganized. The AV node can not regulate all these chaotic currents, and thus leads to the rapid contraction of the ventricles. Since the atria and the ventricles do not beat in a coordinated way, the heart rhythm becomes fast and irregular.

There are many causes of atrial fibrillation. The most common cause is damage to the heart's structure [12]. Other possible causes include:

1. High blood pressure
2. Heart attack
3. Coronary artery disease
4. Exposure to stimulants, such as medications, caffeine, tobacco or alcohol
5. Previous heart surgery
6. Sleep apnea

Besides the items above, atrial fibrillation can also become worse with age. Therefore, as the life expectancy keeps increasing, it is likely that more people will suffer from atrial fibrillation in the future.

Sometimes, people with AF do not show symptoms and only physical examinations can detect the existence of it. More often, though, people with AF may experience the following symptoms [2]:

1. General fatigue
2. Rapid and irregular heartbeat
3. Dizziness
4. Shortness of breath and anxiety
5. Chest pain or pressure

When an irregular heart beat is suspected, the following tests are common options to detect the existence of AF: electrocardiogram (ECG), electrophysiology (EP) study, stress test, and heart monitors [3]. Among these tests, the ECG is a non-invasive way to record the patterns of the heart beat by placing an array of electrodes on the body surface. A typical (schematic) ECG recording of a normal sinus rhythm is given in Figure 1.1.

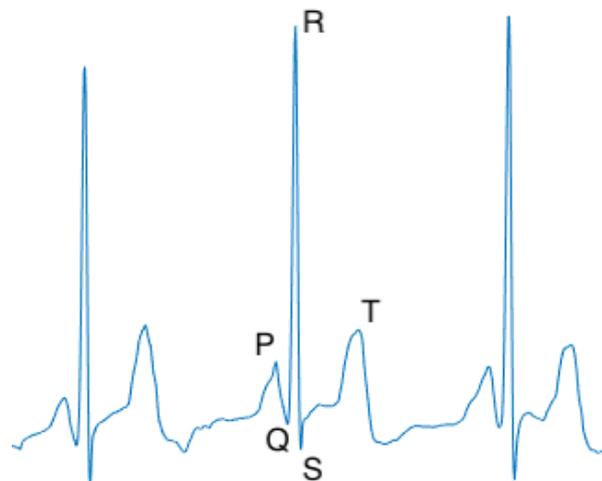


Figure 1.1: Normal sinus rhythm ECG.

Figure 1.1 shows that a normal sinus rhythm ECG signal consists of the P wave, the QRS complex, and the T wave. These waves indicate the sequence of depolarization and repolarization of the atria and ventricles [5]. To be specific, the P wave represents the depolarization of the atria, which is usually 80 to 100 ms in duration. The QRS complex represents the depolarization of the ventricles, whose duration is normally 60 to 100 ms. The T wave represents the repolarization of the ventricles. It is longer in duration than the QRS complex, because the conduction of the repolarization wave is slower than that of the depolarization. When AF is occurring, the absence of P waves and the irregularity of RR intervals (intervals between successive R peaks) are the main characteristics.

1.2. Related work

Most features used to classify AF signals and Normal signals can be categorized into four types: time domain, frequency domain, time-frequency domain, and nonlinear.

1. Time domain features are often calculated from RR intervals, which are the intervals between consecutive R peaks, or P waves. Ye et al. [42] used the local RR interval and the average RR interval as features to characterize the rhythm near a heartbeat, where the local RR interval is the average of ten RR intervals, and the average RR interval is the average RR interval within 5 minutes. Lin et al. [28] normalized the RR intervals by the mean value of all RR intervals within the

same ECG recording, and calculated the average one-minute and twenty-minute RR intervals. Fukunami et al. [20] calculated the P wave duration through an averaging of 200 beats and a significant difference was found between patients with and without paroxysmal atrial fibrillation.

2. Frequency domain features use the Fourier transform to convert ECG signals from the time domain into the frequency domain, where the phase and magnitude information is given per frequency band. Dokur et al. [18] calculated discrete Fourier transform coefficients and found that feature vectors formed by the DFT of the ECG signals scatter in the feature space. Pourbabaee et al. [34] calculated the power spectral density of the ECG signals using the periodogram and lomb estimators. The power spectral density of the whole ECG signal was found suitable for paroxysmal atrial fibrillation prediction.
3. Time-frequency domain features represent ECG signals in both the time and frequency domains concurrently. Guler et al. [21] decomposed the ECG signals into time-frequency representations using the discrete wavelet transform due to its varying window size and an optimal time-frequency resolution in all frequency ranges. Pal et al. [32] proposed an Empirical Mode Decomposition (EMD) based ECG signal enhancement technique to denoise the signals and detect the QRS complex.
4. Nonlinear feature extractor such as recurrence quantification analysis (RQA) [19] and higher-order spectra (HOS) [31] can quantify the concealed characteristics in the ECG signals. The RQA parameters investigate the recurrence of the system and measure the complexities of the ECG signals. The HOS is a spectral representation of moments and cumulants of order greater than two. It is a common parameter that detects and characterizes the nonlinear correlations of the different frequency components of the ECG signals [22].

In addition to the various features discussed above, many researches focused on implementing powerful and novel classifiers and algorithms to solve the problem. For example, in the PhysioNet Computing in Cardiology (CinC) Challenge 2017, whose data is used in this thesis, four groups reached the highest F1 score of 0.83. Datta et al. [17] proposed a two layer binary cascaded approach, which classifies the recordings into one of the two intermediate classes ('normal+others' and 'AF+noisy') before further classification. Hong et al. [23] combined expert features, DNNs (Deep Neural Networks), and centerwave features to train ensemble classifiers. Expert features included features from the statistical area, the signal processing area and the medical area. Centerwave features were proposed by the team which stand for the most representative wave. Zabihi et al. [43] extracted 150 time, frequency, time-frequency, and phase space features out of a total of 491 features. A random forest classifier was then used to classify the selected features. Teijeiro et al. [38] used Tree Gradient Boosting algorithm to evaluate the recordings globally and a Recurrent Neural Network to evaluate each detected heartbeat. The two classifiers are finally combined using a Linear Discriminant Analysis (LDA) classifier.

1.3. Research objectives

This thesis aims to achieve two goals. The first goal is to propose a new feature that can exploit the characteristics of the AF rhythm ECG signals. The second goal is to evaluate the performance of the new feature. The whole procedure can be divided into the following steps:

1. To understand the characteristics of AF rhythm signals and the difficulty of classifying them and normal rhythm signals.
2. To understand the dataset used in this thesis and the potential problems of the data.
3. To understand some state-of-the-art methods and figure out what can be learned from them.
4. To propose a new feature that can extract the characteristics of the AF rhythm.
5. To implement a classifier and learn how to choose the optimal hyper parameters.
6. To evaluate the performance of the classifier and try to improve the results.

1.4. Contribution

The contribution of this thesis is given as follows:

1. A new feature is proposed based on the Poincaré plot. The plot is divided into small bins and the count of bins that are visited by the system at least once is used as the new feature.
2. Mask processing technique is used in this thesis to give each bin a different weight. The new feature uses the distribution of the data from the whole training set in the Poincaré plot, and the relative positions between each bin to the center in the plot as the weight.
3. Autocorrelation functions of both ECG and prediction error are used to generate the Poincaré plot in addition to RR intervals. Peak intervals in the autocorrelation functions are shown to have extra information that can be helpful with the classification.
4. Distribution of the features' value is compared in the box plot among the proposed feature and some other commonly used Poincaré plot based features. The proposed feature is shown to have larger gaps between the AF and Normal classes than other features.
5. The AF and Normal rhythm ECG signals are classified using only Poincaré plot based features by both support vector machine and random forest. The performance of the two classifiers are compared. The results are further improved by means of a feature selection method.
6. The m-lagged analysis of the proposed feature is implemented. It shows that the feature's value for AF rhythm is always higher than that for Normal rhythm, although the gap is decreasing as the increase of the lag of the Poincaré plot.

1.5. Outline

The rest of the thesis is organized as follows. Chapter 2 presents the dataset used in this thesis, and several pre-processing methods that are performed before the feature extraction. Chapter 3 explains concept of the Poincaré plot and some commonly used features based on this plot. A new feature is proposed using mask processing to extract the distribution information and the position information in the Poincaré plot. Autocorrelation function is used to expand the application scope of the Poincaré plot. In chapter 4 the feature selection method and two classifiers including support vector machine and random forest are introduced. Chapter 5 compares the results of different features and compares the results of the two classifiers. The performance is further improved by the feature selection method. Chapter 6 discusses the conclusion drawn from the results, and describes some future directions.

2

Data and pre-processing

In this chapter, the data and the pre-processing steps are discussed. Section 2.1 describes the Physionet/Computing in Cardiology (CinC) challenge 2017 dataset. In Section 2.2, R peak detection algorithm, which combines an envelope-based method and the modified Pan-Tompkins algorithm, is explained. In Section 2.3, some pre-processing steps are applied, including lead inversion detection, normalization, and outlier removal.

2.1. Data

The dataset used in this thesis project comes from the PhysioNet/Computing in Cardiology (CinC) Challenge 2017, which focused on classifying AF from normal, noisy and other rhythms in ECG recordings [15]. The challenge provided 12,186 recordings in total, of which 8,528 recordings are in the public training set and 3,658 recordings are in the hidden test set. These recordings were obtained from AliveCor’s single-channel ECG device. Due to the inaccessibility of the hidden testing set, only the public training set was used in this thesis project. The signal length of the training set varies from 9 s to 61 s, with a mean signal length of 32.5 s. The sampling frequency is 300 Hz. The recording labels comply with the latest version in the challenge, where Normal rhythm constitutes more than half of the whole training set, and Noisy rhythm only has less than 300 recordings. More details on the final version of the labels can be seen in Table 2.1.

Table 2.1: Data profile for the training set

Label	Number of recordings	Percentage (%)
Normal	5076	59.5
AF	758	8.9
Noisy	279	3.3
Other	2415	28.3
Total	8528	100

Some examples of the ECG recordings for Normal, AF, Other, and Noisy rhythms are shown in Figure 2.1 from top to bottom, respectively. In this thesis project, only Normal and AF rhythms were used.

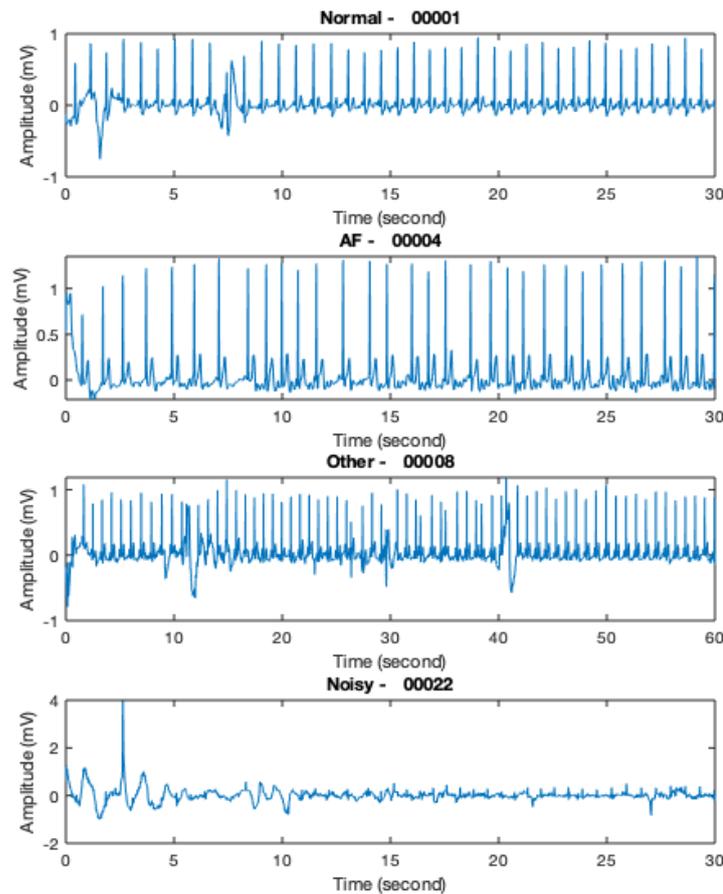


Figure 2.1: Examples of ECG recordings. The labels and the indices of the recordings are given above each figure (For example, "Normal - 00001" means the first recording in the dataset coming from the normal class).

2.2. R peak detection

Many features are based on the analysis of the heart rate variability (HRV), which requires the accurate detection of R peaks. This project used an available matlab toolbox 'R-DECO' [30], which combines an envelope-based procedure and the modified Pan–Tompkins algorithm, to detect the R peaks. The algorithm consists of three steps:

- 1) QRS complex enhancement: The upper (U) and lower (L) envelopes of the ECG signal are first calculated by the secant method on $150ms$. They are then subtracted from each other to obtain a flattened ECG signal F : $F = U - L$. This emphasizes the QRS complex and flattens the remaining of the ECG signal. This is shown in the Figure 2.2, where the grey solid line represents the ECG signal.

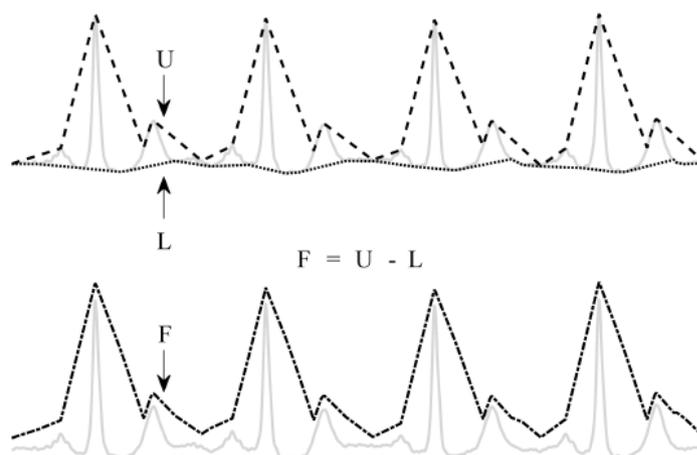


Figure 2.2: Visual explanation of enveloping procedure (Figure from [30]).

2) Searching of R peaks:

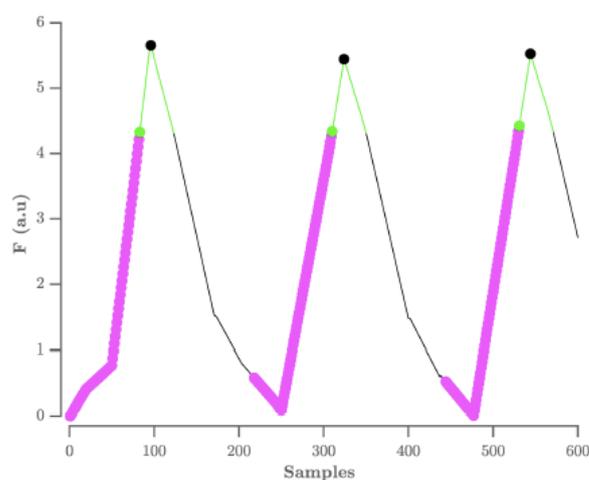


Figure 2.3: R peaks selection (Figure from [30]). The purple circles indicate the upward slopes, the green line indicates the searching window, and the black circles indicate the R peaks.

As shown in Figure 2.3, at first the upward slopes are identified (purple circles) as the samples lower than the samples 80 ms ahead of them. Then the slopes shorter than 80 ms are excluded to avoid small peaks. Lastly, the peaks are selected within a window (green line) starting at the end of the upward slope. The peaks are defined using the adaptive thresholding procedure of the Pan–Tompkins algorithm.

- 3) Post-processing: In this step, the original ECG signal is used to find the exact position of the R peaks by searching within 50 ms from the R peaks identified in step 2. The goal of this search-back procedure is to avoid peak shift due to S waves.

2.3. Pre-processing

In this section, pre-processing steps including lead inversion detection, normalization of the RR intervals and outlier removal are implemented on the raw ECG recordings to reduce the influence of the irrelevant factors.

2.3.1. Lead inversion detection

Many ECG recordings in the dataset were found to be inverted, which is probably because the electrodes were misplaced or held in the wrong orientation. In the inverted signal, the “R peaks” detected

by the algorithm could be in fact a Q wave or an S wave. Without proper inspection, this will result in wrong classification.

One way of checking whether an ECG signal is inverted is to check whether the value of the T wave is positive or not. However, this requires an additional detection procedure for the T wave, which is not needed in this thesis project. Therefore, an extra R peak detection was implemented after the ECG signals were inverted. If the average peak height of the inverted signal is higher than the original one, then the original signal is considered to be inverted and requires re-inversion.

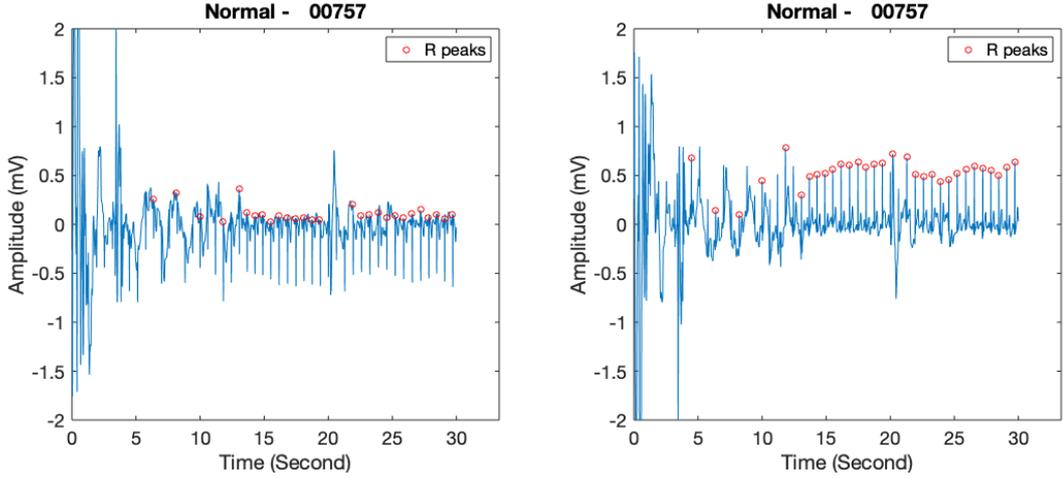


Figure 2.4: An example showing an original ECG signal (left) and its inversion (right).

For example, the left figure in Figure 2.4 shows an inverted ECG signal. But the R peak detection algorithm can still result in a series of positive “R peaks”. The average peak height is 0.37 mV. In the figure on the right side, the whole signal is inverted and the R peak detection algorithm is implemented again. This time, the average peak height is 0.53 mV. So the lead inversion detection algorithm will choose the signal in the right figure in Figure 2.4.

2.3.2. Normalization of the RR intervals

The heart rate can be calculated from the RR intervals. This can change a lot among people. A healthy heart rate varies from 60 to 100 beats per minute when in rest [7]. The maximum of the heart rate during exercise could be roughly calculated as 220 heart beats per minute minus the age of that person, according to the American Heart Association (AHA). So, for example, the maximum heart rate for a 40 year-old man is around 180 heart beats per minute. Therefore, both the status and the age of a person can influence his heart rate.

Different heart rates can impact the performance of heart rate variability (HRV) based features. HRV is the variation in the time interval between consecutive heartbeats. Chapter 1 already mentioned that AF causes the heart rate to be aperiodic. When checking the irregularity of RR intervals of AF signals, the difference between consecutive RR intervals of signals with a higher heart rate might not be as obvious as that of those signals with a lower heart rate. In other words, a person with a fast heart rate has shorter RR intervals. When suffering from AF, the changes of RR intervals are relatively small, and thus makes it difficult to separate from a Normal rhythm.

In order to get rid of the influence of the heart rate when extracting HRV based features, the heart rate of all recordings needs to be normalized to the same scale at first. In this thesis, this is achieved by dividing all the RR intervals by the average value of RR intervals of that recording:

$$RR_{Normalized} = \frac{RR}{mean(RR)}. \quad (2.1)$$

This means the unit of the normalized RR intervals is no longer in milliseconds, but a ratio. Hence two recordings with different heart rate will have the same mean RR interval value (equal to one), and are thus comparable with each other.

Figure 2.5 shows examples of the tachograms for the normalized RR intervals and the corresponding ECG recordings, with Normal rhythm on the left and AF rhythm on the right. In the tachogram, the normalized RR intervals are plotted on the y-axis against the time at which the R peak occurs on the x-axis. Compared with the Normal class, the AF class shows a higher fluctuation for the normalized RR intervals along time in the tachogram. It should be noted that part of the reasons that lead to the high value of the normalized RR intervals around 10 seconds for the AF class is because of the failure of the R peak detection, which can be seen in the corresponding ECG recording. To partially reduce the influence of the failure of the R peak detection algorithm, some outliers are removed in the next subsection.

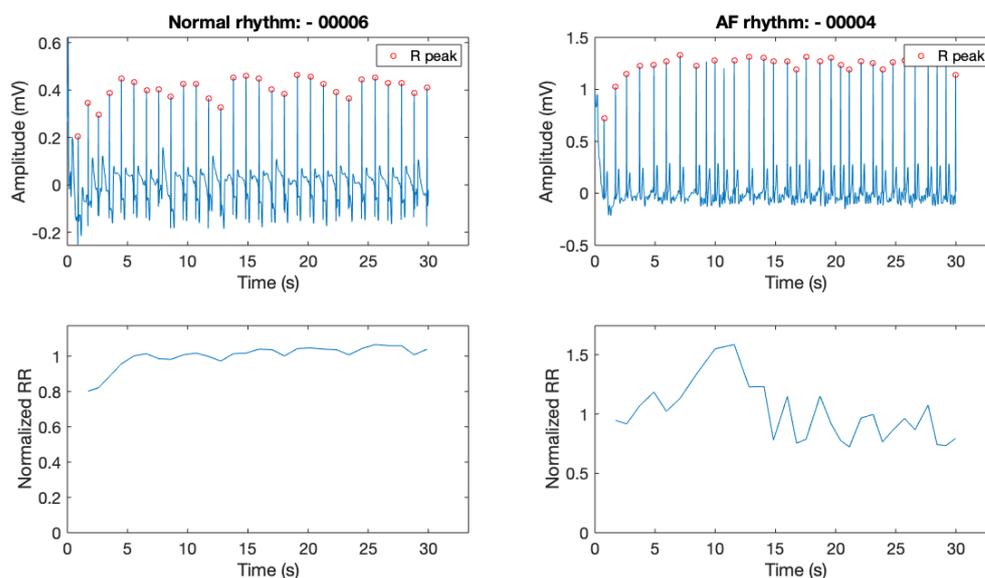


Figure 2.5: Examples of the tachograms (the second row) and the ECG recordings for the two classes (the first row).

2.3.3. Outlier removal

After the normalized RR intervals are calculated, most intervals are centered around the average (which is thus equal to one), while a number of recordings are found to have outliers larger than 2, sometimes even up to 5. This is not likely to be caused by AF, because during AF periods, the heart is contracting faster, leading to a smaller RR interval. After a careful check, some outliers are found to be caused by artifacts. An example is shown in Figure 2.6, with the plot of the signal on the left and the histogram of normalized RR intervals on the right. It can be seen from the left figure that around 50 seconds, the signal is rather different from other parts of the signal, which results in a RR interval of more than four times the average interval.

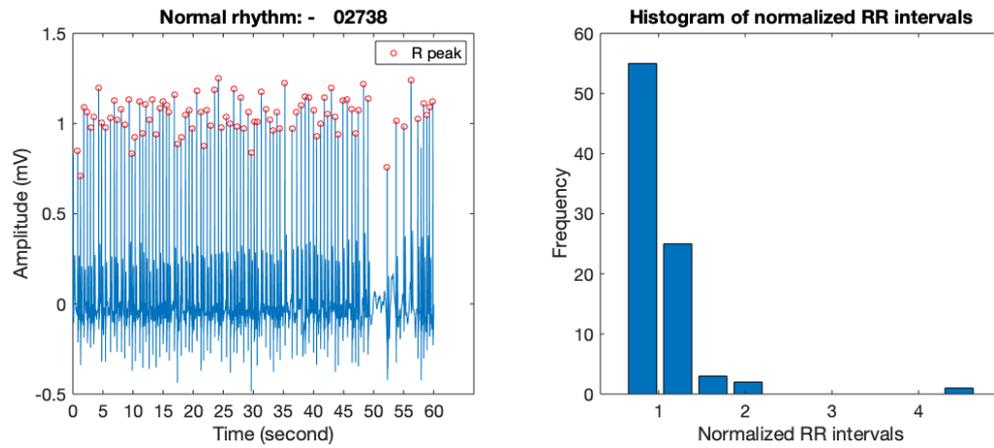


Figure 2.6: Outlier due to artifacts.

Other outliers are found to be caused by failure of the R peak detection algorithm. An example is shown in Figure 2.7. It can be seen in the left figure that between 40 to 45 seconds, three consecutive R peaks are neglected by the detection algorithm (marked by a large red ellipse). This leads to an RR interval with a length of more than three times the average interval.

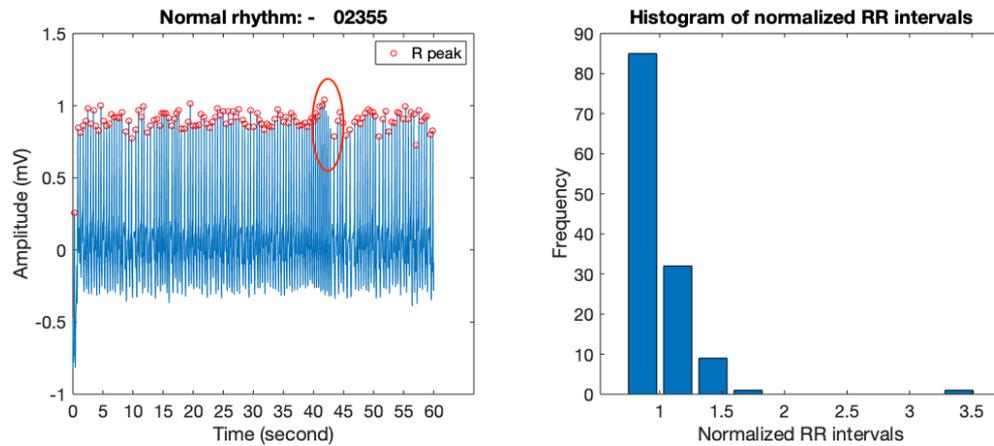


Figure 2.7: Outlier due to failure of R peak detection.

These outliers do not contribute to the correct classification of AF and Normal ECG signals. Thus, in this thesis, any RR interval larger than twice the average interval is removed.

3

Feature extraction

In this chapter, the procedure for feature extraction is introduced. In Section 3.1, the concept of the Poincaré plot is introduced. In Section 3.2, some Poincaré plot based features are explained. In Section 3.3, a new Poincaré plot based feature is proposed. In Section 3.4, the autocorrelation function is used to generate an alternative Poincaré plot.

3.1. Poincaré plot

The Poincaré plot is a type of recurrence plot used to evaluate the self-similarity nature of states. It has shown to be a useful graphical representation to analyze the nonlinear aspects of the heart rate dynamics [14]. It presents the correlation (by calculating the Pearson's correlation coefficient) between two consecutive data points in a time series, with a greater correlation between the present and past implying the presence of memory in the time series [36]. The Poincaré plot can be evaluated qualitatively by categorizing the shape into a fan or comet as normal, or a random pattern with asymmetrical RR interval clusters as abnormal. It can also be evaluated quantitatively to provide an estimation of short-term and long-term variability of a time series (by calculating the standard deviation perpendicular and parallel to the line of identity) [24]. The quantitative analysis is based on the notion of different temporal effects of changes in the vagal and sympathetic modulation of the heart rate on the subsequent RR intervals, which does not require the data to be stationary [40].

Let X_n ($n = 1, 2, \dots, M$) be a sequence of data points, the lag-1 Poincaré plot presents X_{n+1} on the y-axis as a function of X_n on the x-axis. RR intervals are widely investigated as the data sequence in the Poincaré plot, where the values of a pair of consecutive RR intervals represent a point in the plot. The idea of using consecutive points is based on the assumption that the current point significantly influences the next point [26]. The plot provides beat-to-beat information of the patterns that are generated by the nonlinear processes and thus can not be detected by the standard time-domain analysis of the HRV signal. The variability of the RR intervals increases from the onset till the end of AF. Besides, baseline wandering and noise are more likely to contaminate the morphology of the fibrillatory waves than the HRV [33], which makes visual inspection of the patterns of HRV a useful method to detect AF.

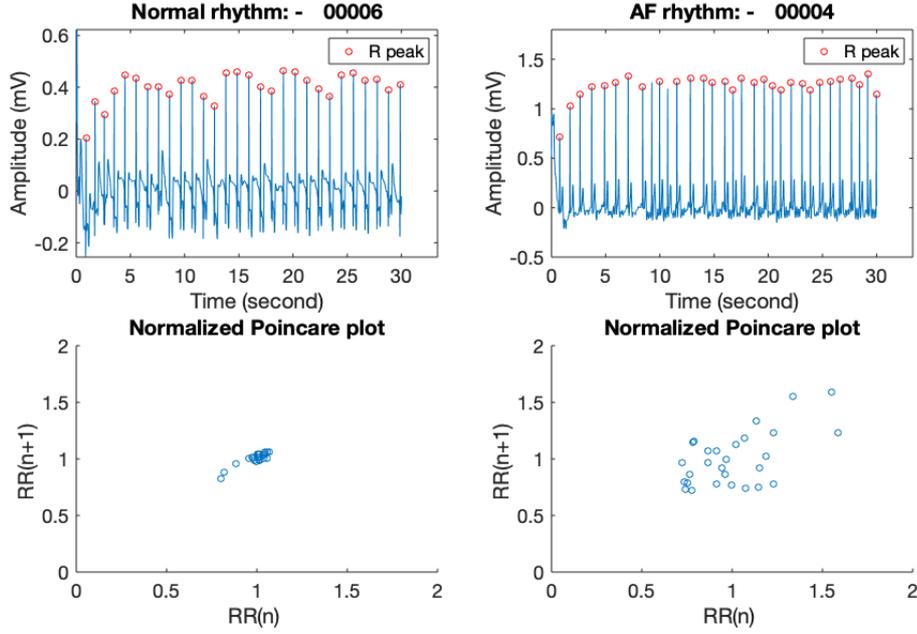


Figure 3.1: The normalized Poincaré plots for a normal ECG signal (left) and an AF ECG signal (right).

Figure 3.1 illustrates interval $RR(n+1)$ as a function of the previous interval $RR(n)$ to show the difference between Normal rhythm and AF rhythm. The figures on the top left and bottom left show a Normal ECG recording and the corresponding Poincaré plot, respectively. Most points can be found around the center in the Poincaré plot, which implies the corresponding RR intervals have a low variance. Still, these points show a higher variance along the identity line than the direction perpendicular to it, which means the signal has a relatively large long-term correlation and a small short-term correlation.

The figures on the top right and bottom right, on the other hand, show an AF ECG recording and the corresponding Poincaré plot, respectively. The Poincaré plot on the bottom right has a higher dispersion than that of the normal signal, which means the corresponding RR intervals are more irregular. To be specific, the variances along both the identity line and the direction perpendicular to it are higher than those in the case of Normal rhythm recordings, meaning the short-term correlation and long-term correlation of an AF rhythm signal are both large. Therefore, this provides some promising ways to differentiate an AF signal from a Normal signal, as can be seen in the next section.

3.2. Commonly used Poincaré plot based features

The previous section explained the basic concept of the Poincaré plot. This section introduces some commonly used features based on the Poincaré plot together with a proposed modified Poincaré plot feature F_7 which will be explained in detail in the section 3.3. These features will be calculated directly from the RR intervals in the ECG signals at first, but this thesis also investigates the scenario of calculating these features from the peak intervals in the autocorrelation of the ECG signal and of the prediction error. The latter two will be explained in the section 3.4. Therefore, 21 features in total are used in this thesis.

All the features used in this thesis can be seen in Table 3.1. All in all, features F_1 to F_3 reflect the standard deviation of the data. Features F_4 and F_5 capture the temporal structure of the plot. Feature F_6 shows the sparseness of the data in the plot. Feature F_7 uses the distribution and position information of the data in the plot as weight to each bin in the feature F_6 . The details of each feature will be explained in the following subsections.

Table 3.1: All the extracted features

Annotation	Features	Explanation
F_1	SD1	Standard deviation along the identity line in the Poincaré plot
F_2	SD2	Standard deviation along the direction perpendicular to the identity line
F_3	SD1/SD2	Ratio of SD1 and SD2
F_4	Mean stepping increment [33]	Mean distance between the consecutive points in the Poincaré plot
F_5	Complex correlation measure (CCM) [26]	Mean area of the triangles composed of the consecutive three points in the Poincaré plot
F_6	Unweighted area of bins	Count of bins in the Poincaré plot that are visited by the system at least once
F_7	Weighted area of bins	Giving weights to the bins in feature F_6

3.2.1. Ellipse-fitting technique

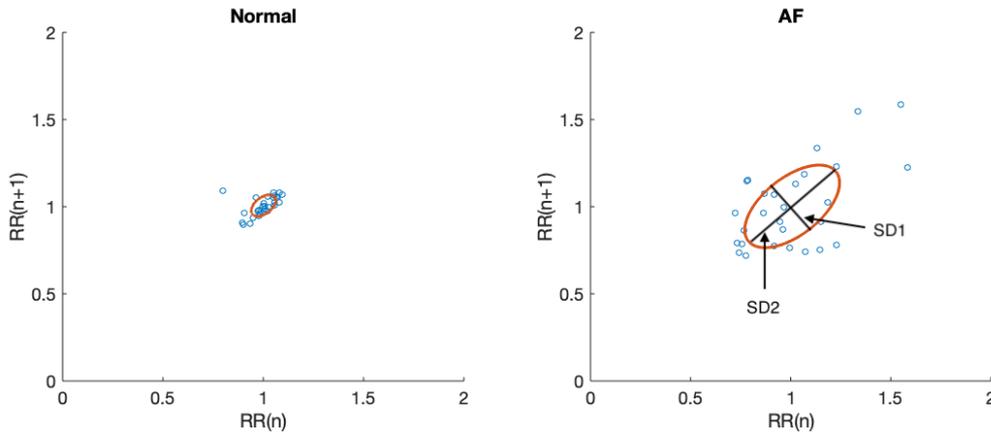


Figure 3.2: Popular Poincaré plot features: SD1 and SD2. The Poincaré plot is less scattered for the Normal rhythm compared with the AF rhythm.

Figure 3.2 shows a popular way to quantitatively analyze the Poincaré plot [40]: fitting an ellipse to the plot. The center of the ellipse is at the center of the plot, the major axis is along the identity line, and the minor axis is perpendicular to the major axis. The standard deviation of the data along the minor axis, denoted as SD1, represents the short-term variability of the data, which reflects the parasympathetic activity and relates mainly to the effects of respiration on vagal drive [29]. The standard deviation of the data along the major axis, denoted as SD2, represents the long-term variability of the data, which reflects the sympathetic modulation and relates to other heart rate changes, including those associated with sympathetic oscillations, baroreflex loop, thermoregulation, and fluctuations in humoral factors. If the heart rhythm is regular, the points in the plot will be populated around the line of identity.

The line of SD2 and SD1 can be expressed as:

$$SD2 : RR(n+1) = RR(n) \quad (3.1)$$

$$SD1 : RR(n+1) = -RR(n) + 2\overline{RR(n)} \quad (3.2)$$

where $\overline{RR(n)}$ is the mean value of RR intervals. The distance of point (x_0, y_0) from a line $Ax+By+C=0$

can be expressed as:

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}} \quad (3.3)$$

Let d_{n1} and d_{n2} be the distance from the n th point ($RR(n), RR(n+1)$) in the plot to the major and minor axis, respectively:

$$d_{n1} = \frac{|RR(n) - RR(n+1)|}{\sqrt{2}} \quad (3.4)$$

$$d_{n2} = \frac{|RR(n) + RR(n+1) - 2\overline{RR(n)}|}{\sqrt{2}}. \quad (3.5)$$

SD1 and SD2 can thus be expressed as

$$SD1 = \sqrt{\frac{1}{M-1} \sum_{n=1}^{M-1} d_{n1}^2} \quad (3.6)$$

$$SD2 = \sqrt{\frac{1}{M-1} \sum_{n=1}^{M-1} d_{n2}^2} \quad (3.7)$$

where M is the number of RR intervals.

Let SDNN be the standard deviation of the RR intervals:

$$SDNN = \sqrt{\frac{1}{M} \sum_{n=1}^M (RR(n) - \overline{RR(n)})^2} \quad (3.8)$$

and SDSD be the standard deviation of the successive differences of the RR intervals:

$$SDSD = \sqrt{\frac{1}{M-1} \sum_{n=1}^{M-1} (\delta RR(n) - \overline{\delta RR(n)})^2} \quad (3.9)$$

where

$$\delta RR(n) = RR(n) - RR(n+1) \quad (3.10)$$

and $\overline{\delta RR(n)}$ is the mean value of $\delta RR(n)$. Note that $\overline{\delta RR(n)} = E[RR(n)] - E[RR(n+1)] = 0$ for stationary intervals. Thus, SDSD is statistically equivalent to the root mean square of the successive differences of the RR intervals. According to [14], these descriptors, SD1 and SD2, are still related to the time domain statistics of the HRV, which are SDSD and SDNN:

$$SD1 = \frac{1}{\sqrt{2}} SDSD \quad (3.11)$$

$$SD2 = \sqrt{2SDNN^2 - \frac{1}{2}SDSD^2} \quad (3.12)$$

$$SD1^2 + SD2^2 = 2SDNN^2 \quad (3.13)$$

3.2.2. Temporal based features

The limitation of the features F_1 to F_3 is that they reveal the global variability of the RR intervals, and fails to capture the nonlinear temporal structure of the plot quantitatively. Further, when there are multiple clusters in the plot due to complex dynamic behaviors, the above features yields mixed results because they rely on the existence of a single cluster or a defined pattern [26]. On the other hand, features F_4

and F_5 aim to investigate the temporal information of the time series, which is distinguishing between Normal and AF, and these features are based on the correlation between consecutive two and three points in the plot, respectively.

Feature F_4 in Table 3.1 calculates the mean distance of the consecutive points in the Poincaré plot divided by mean value of intervals:

$$F_4 = \frac{\frac{1}{M-2} \sum_{n=1}^{M-2} \sqrt{(RR(n) - RR(n+1))^2 + (RR(n+1) - RR(n+2))^2}}{\frac{1}{M} \sum_{n=1}^M RR(n)}. \quad (3.14)$$

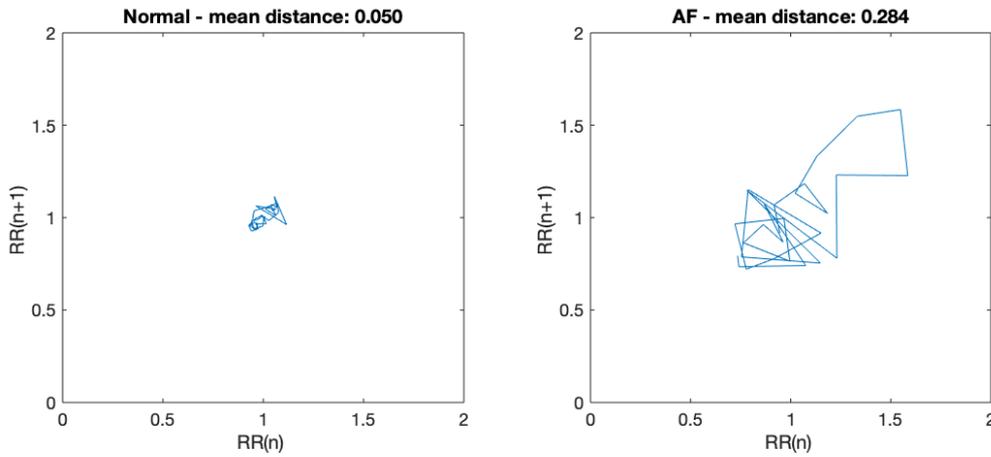


Figure 3.3: Schematic of feature F_4 .

Figure 3.3 illustrates the feature F_4 for a Normal signal and an AF signal, respectively. Since the points for Normal rhythm are clustering in the center, the mean distance between points is only 0.050. This means the RR intervals are almost uniformly distributed. The points for AF rhythm do not have a certain pattern and disperse highly in the plot. This indicates that the RR intervals are statistically independent from each other, except a small correlation between the immediate subsequent points [33]. This leads to a higher mean distance between points: 0.284, which is more than five times of that of a Normal rhythm.

Feature F_5 in Table 3.1 calculates the mean area of the triangles composed of the consecutive three points in the Poincaré plot divided by the area of the fitting ellipse. Let $||S(i)||$ be the area of the i th triangle, and $\pi \times SD1 \times SD2$ represent the area of the fitted ellipse over the Poincaré plot, F_5 can be expressed as:

$$F_5 = \frac{1}{\pi \times SD1 \times SD2 \times (M-2)} \sum_{i=1}^{M-2} ||S(i)||. \quad (3.15)$$

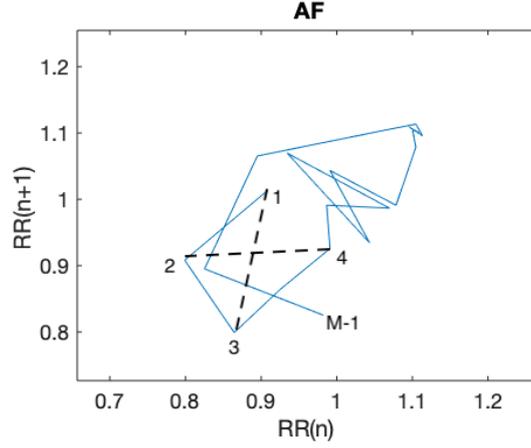


Figure 3.4: Schematic of feature F_5 .

Figure 3.4 illustrates about how to define triangles in feature F_5 for an AF signal, where the triangles are composed of the points 123, 234, ..., (M-3)(M-2)(M-1), respectively. In this way, not only the distance between points are contributing to the feature, but also the relative direction among the consecutive three points will impact the final value. Using the same two examples as in Figure 3.3, $F_5 = 0.1849$ for Normal rhythm, and $F_5 = 0.232$ for AF rhythm. This feature again shows a higher value for AF rhythm and a lower value for Normal rhythm.

3.2.3. Count of bins in the Poincaré plot

The feature F_6 measures the sparseness of the Poincaré plot. First, the plot is quantized into Q levels along each dimension, as can be seen in Figure 3.5. The bin is blue if at least one point is located inside the bin, otherwise it is white.

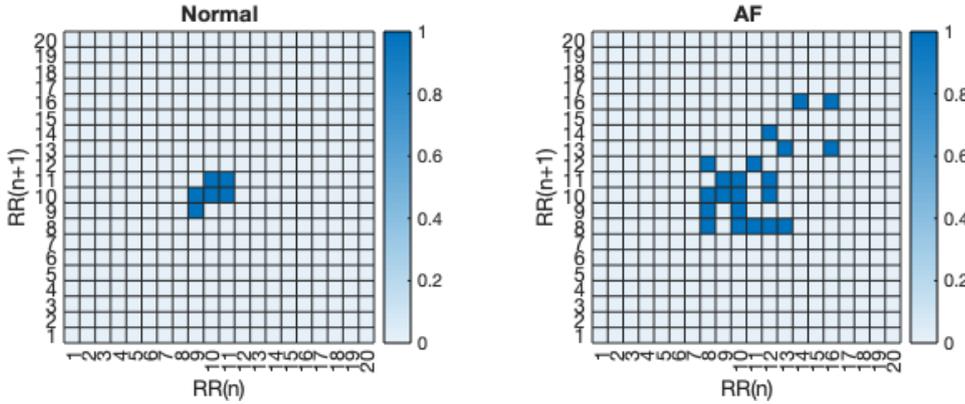


Figure 3.5: Quantization of Poincaré plot, with normal sinus rhythm on the left and AF on the right.

This can then be easily expressed as a $Q \times Q$ matrix \mathbf{R} :

$$\mathbf{R} = \begin{bmatrix} R_{1,1} & \cdots & R_{1,Q} \\ \vdots & \ddots & \vdots \\ R_{Q,1} & \cdots & R_{Q,Q} \end{bmatrix} \quad (3.16)$$

where:

$$R_{i,j} = \begin{cases} 1, & \text{If a particular } (RR_i, RR_j) \text{ combination occurred in the Poincaré plot at least once,} \\ 0, & \text{otherwise,} \end{cases} \quad (3.17)$$

where i, j denote the row and column of the matrix, respectively. It can be seen from Figure 3.5 that for normal sinus rhythm signals, less states occurred due to its periodic characteristics. This results in

less non-zero entries in \mathbf{R} and most of them are populating around the center of \mathbf{R} , while for AF signal the opposite is the case.

In order to encode information contained by patterns in the Poincaré plot, let F_6 be the feature showing the regularity characteristic of RR intervals, which can be expressed as the number of states visited by the system at least once:

$$F_6 = \sum_{i,j=1}^Q R_{i,j}. \quad (3.18)$$

This feature measures the density of the matrix, with high values for AF signals and low values for Normal signals. In Figure 3.5, $F_6 = 6$ for normal sinus rhythm on the left, while $F_6 = 21$ for AF on the right.

However, the separable case as in Figure 3.5 does not occur all the time. For example, some Normal signals occasionally contain outlier bins (i.e., states), which is possibly due to noise. This would have been noticed by manual check because the outlier bins occur at improbable places. In other cases, some AF signals might contain only few clusterings in the plot and have fewer bins visited by the system than a typical AF signal, which leads to a low value of feature F_6 . This could have been corrected because the clusterings of the points are far from each other and thus the signal is not likely to be normal rhythm. In these cases, the feature F_6 assumes each bin in the Poincaré plot has the same weight contributing to the final value of the feature. However, this may lose some crucial information concerning the distribution of data that each bin conveys, and the relative position that each bin locates compared with other bins, as mentioned in the above examples. Therefore, the concept of mask processing is introduced in the next section to solve the problem.

3.3. Weighted count of bins in the Poincaré plot

3.3.1. Mask processing

Mask processing technique has been commonly applied to filter the image. A mask is a matrix whose values are called weights. For a 2D mask, it gives weights to the entries of the input matrix by element-wise multiplying with the input matrix and summing all the results to get one value at the corresponding place, as the red part shows in Figure 3.6. Then the mask slides across the whole input matrix to obtain the output matrix.

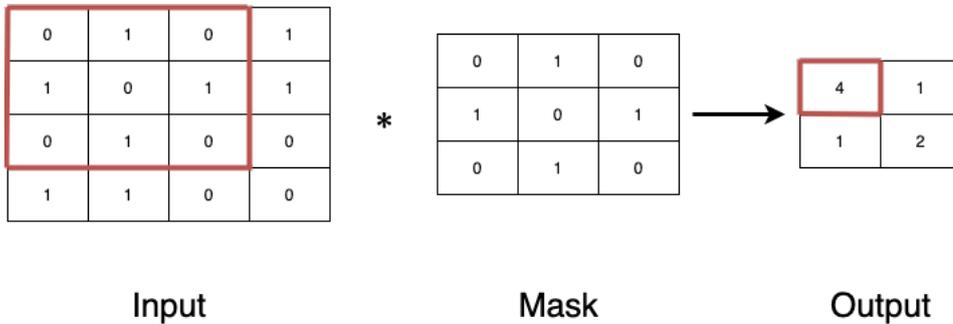


Figure 3.6: Diagram of the mask processing.

By doing so, certain pattern in the input matrix that is similar to the mask can be extracted by the filter and the element-wise multiplication and summation operation results in a high value. A pattern in the input matrix that is different from the mask does not activate the filter and leads to a low value [6], as the low value in the output matrix shows in Figure 3.6.

The mask used in this thesis is a two-dimensional $Q \times Q$ matrix \mathbf{W} :

$$\mathbf{W} = \begin{bmatrix} W_{1,1} & \cdots & W_{1,Q} \\ \vdots & \ddots & \vdots \\ W_{Q,1} & \cdots & W_{Q,Q} \end{bmatrix} \quad (3.19)$$

The operation that contains element-wise multiplication and summation between \mathbf{R} and \mathbf{W} is defined

by:

$$\mathbf{R} * \mathbf{W} = \begin{bmatrix} R_{1,1} & \cdots & R_{1,Q} \\ \vdots & \ddots & \vdots \\ R_{Q,1} & \cdots & R_{Q,Q} \end{bmatrix} * \begin{bmatrix} W_{1,1} & \cdots & W_{1,Q} \\ \vdots & \ddots & \vdots \\ W_{Q,1} & \cdots & W_{Q,Q} \end{bmatrix} \quad (3.20)$$

$$= \sum_{i=1}^Q \sum_{j=1}^Q R_{i,j} W_{i,j} \quad (3.21)$$

It should be noted that the masks used in this thesis have the same size as the input matrix, while in image processing the mask is much smaller than the input image. This is because an image might have hundreds of pixels along one dimension, and a mask with the size of 3×3 or 5×5 extracts the local features from everywhere of the image and helps reduce the storage of the mask parameters. But this is not the case for the matrix \mathbf{R} which indicates if the state combination (RR_i, RR_j) of the Poincaré plot is visited by the system. A matrix \mathbf{W} with the same size as a matrix \mathbf{R} does not cause the problem of storing too many parameters and the mask aims to extract the features globally. So it only implements the element-wise multiplication and summation operations without the sliding part as it is for image processing.

Clearly, different masks \mathbf{W} can extract different features from \mathbf{R} . This thesis proposes two kinds of masks, $\mathbf{W1}$ and $\mathbf{W2}$, to detect AF pattern in matrix \mathbf{R} , which will be explained in the section 3.3.2 and 3.3.3.

3.3.2. Distribution of bins in the Poincaré plot

The first mask $\mathbf{W1}$ gives weights to the elements in the matrix \mathbf{R} and aims to obtain a high value for the AF class and a low value for the Normal class after the mask processing. This is done by combining the distributions of the two classes in the Poincaré plot and the probability of a new data belonging to the two classes as the value of weight in the mask $\mathbf{W1}$.

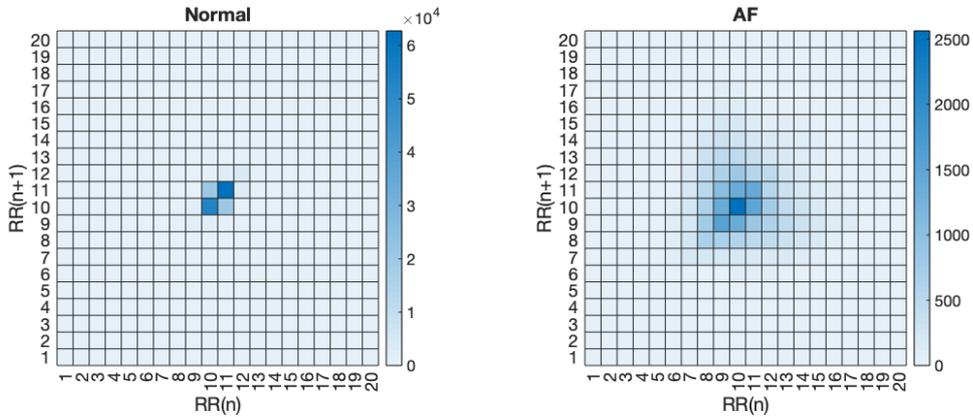


Figure 3.7: 2D histogram of Normal (left) and AF (right) sinus rhythm showing the times of each state (RR_i, RR_j) that was visited by the system.

Figure 3.7 is a 2D histogram of the Poincaré plot of all Normal and AF sinus rhythm recordings for the training data. It shows an approximate distribution of Normal and AF sinus rhythms in the Poincaré plot, with the value of each bin representing the number of times each state (RR_i, RR_j) visited by the system. It can be seen that for Normal signals, the system is more likely to visit states closer to the center of the histogram. This means stable heart rate and little variability between beats no matter short-term or long-term. For AF signals, the histogram shows a higher dispersion, meaning the system is experiencing a larger beat-to-beat fluctuation. Therefore, both short-term and long-term variabilities between RR intervals increase, while the short-term variance witnesses a larger rise.

Let \mathbf{N} and \mathbf{A} be two $Q \times Q$ matrices, which stand for the histogram of Normal and AF sinus rhythm

in Figure 3.7 respectively:

$$\mathbf{N} = \begin{bmatrix} N_{1,1} & \cdots & N_{1,Q} \\ \vdots & \ddots & \vdots \\ N_{Q,1} & \cdots & N_{Q,Q} \end{bmatrix} \quad (3.22)$$

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,Q} \\ \vdots & \ddots & \vdots \\ A_{Q,1} & \cdots & A_{Q,Q} \end{bmatrix} \quad (3.23)$$

where $N_{i,j}$ is the number of times that a particular (RR_i, RR_j) combination occurred in the Poincaré plot for all Normal sinus rhythm recordings in the training set, and similar for $A_{i,j}$. The matrices are then scaled to the range $[0, 1]$ by:

$$\mathbf{N} = \frac{\mathbf{N} - \min(\mathbf{N})}{\max(\mathbf{N}) - \min(\mathbf{N})} \quad (3.24)$$

$$\mathbf{A} = \frac{\mathbf{A} - \min(\mathbf{A})}{\max(\mathbf{A}) - \min(\mathbf{A})} \quad (3.25)$$

to make sure distributions of Normal and AF signal have the same weight in the mask $\mathbf{W1}$.

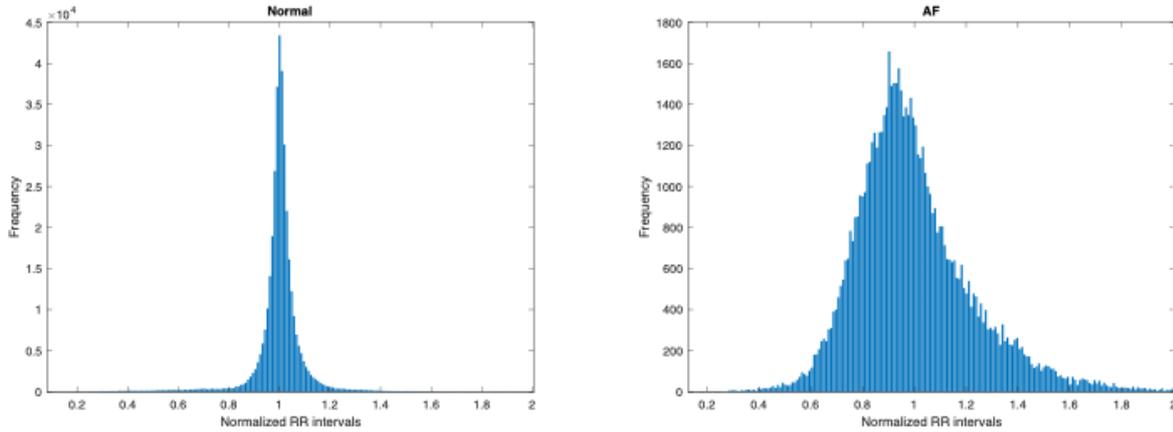


Figure 3.8: Frequency distribution of RR intervals .

Figure 3.8 shows the frequency distribution of RR intervals for all the Normal ECG signals (left) and AF ECG signals (right). It can be seen that $RR(n)$ of Normal signals have a lower variance than those of AF signals.

Before computing the probability of a new data coming from the two classes, whether the distributions in Figure 3.8 come from a distribution in the normal family needs to be checked at first. This is done by performing a hypothesis test called Lilliefors test at first [9]. The null hypothesis is that the sample comes from a distribution in the normal family, against the alternative that it does not come from such a distribution. The test statistics is

$$O_1 = \max_x (|H_1(x) - G(x)|) \quad (3.26)$$

where $H_1(x)$ is the empirical distribution function of the sample data, and $G(x)$ is the cumulative distribution function of the normal distribution with estimated parameters equal to the sample parameters. The result turned out to reject the null hypothesis, so that neither of the distributions is Gaussian. This can be easily seen in Figure 3.8 for AF ECG signals which have an asymmetric distribution around value one and thus should not be assumed to be Gaussian.

Since the distributions are not supposed to be assumed Gaussian, the probabilities are calculated using Kolmogorov–Smirnov test, which is a nonparametric hypothesis test that evaluates the difference between the empirical distribution functions of the distributions of two samples. The null hypothesis is

that the two samples are from the same continuous distribution, while the alternative hypothesis is that the two samples are from different continuous distribution. The test statistic is

$$O_2 = \max_x (|H_1(x) - H_2(x)|) \quad (3.27)$$

where $H_1(x)$ and $H_2(x)$ are the empirical distribution functions of two samples.

If the null hypothesis is assumed to be correct, the probability that a higher test statistic is observed is denoted as p-value. In other words, p-value can be regarded as the probability that the null hypothesis is accepted. Commonly, if $p < 0.05$, the two samples can be assumed to come from different distributions. In this paper, the classical one-dimensional Kolmogorov–Smirnov test is used for calculating the similarity between RR intervals. First, the hypothesis test is made between the RR intervals of a new data and all the Normal data. p-value of Normal, denoted as p_n , is obtained which is proportional to the probability that this new data comes from Normal class. Then, the hypothesis between the same new data and all the AF data is made, and p-value of AF, denoted as p_a , is obtained which is proportional to the probability that this new data comes from AF class. Then these two p-values are normalized:

$$\tilde{p}_N = \frac{p_n}{p_n + p_a} \quad (3.28)$$

$$\tilde{p}_A = \frac{p_a}{p_n + p_a} \quad (3.29)$$

where \tilde{p}_N and \tilde{p}_A are the normalized p-values that indicate the probabilities of a particular realization generated by Normal rhythm distribution and AF rhythm distribution respectively, and the summation of these values equals to one:

$$\tilde{p}_N + \tilde{p}_A = 1 \quad (3.30)$$

Now that every term needed for calculating $\mathbf{W1}$ is available, the equation is given as follows. Let the first mask $\mathbf{W1}$ be a $Q \times Q$ matrix, which extracts features based on the combination of the distribution of the Poincaré plot for the two classes and the probability of a new data generated by the two classes:

$$\mathbf{W1} = \tilde{p}_N \mathbf{N} + \tilde{p}_A \mathbf{A}. \quad (3.31)$$

In equation (3.31), \tilde{p}_N and \tilde{p}_A are proportional to the probability of a new data coming from the class Normal and AF, respectively. Each element in the matrices \mathbf{N} and \mathbf{A} is proportional to the probability that the corresponding state in the Poincaré plot is visited by the system, given the class is Normal or AF. Therefore, equation (3.31) is approximately equal to:

$$\mathbf{W1} \approx \rho(N)\rho(\mathbf{R}|N) + \rho(A)\rho(\mathbf{R}|A) \quad (3.32)$$

where $\rho(N)$ and $\rho(A)$ are the marginal probabilities, and $\rho(\mathbf{R}|N)$ and $\rho(\mathbf{R}|A)$ are the conditional probabilities. According to the law of total probability,

$$\rho(\mathbf{R}) = \rho(N)\rho(\mathbf{R}|N) + \rho(A)\rho(\mathbf{R}|A) \quad (3.33)$$

where $\rho(\mathbf{R})$ is the probability that each element in the matrix \mathbf{R} is visited by the system. Therefore, what equation (3.31) does is approximating the probability of each state in the Poincaré plot being visited by the system for a new data:

$$\mathbf{W1} \approx \rho(\mathbf{R}). \quad (3.34)$$

After the mask $\mathbf{W1}$ is obtained, the element-wise multiplication and summation between \mathbf{R} and $\mathbf{W1}$ (denoted as $\mathbf{R} * \mathbf{W1}$) is given as:

$$\mathbf{R} * \mathbf{W1} = \begin{bmatrix} R_{1,1} & \cdots & R_{1,q} \\ \vdots & \ddots & \vdots \\ R_{q,1} & \cdots & R_{q,q} \end{bmatrix} * \begin{bmatrix} W1_{1,1} & \cdots & W1_{1,q} \\ \vdots & \ddots & \vdots \\ W1_{q,1} & \cdots & W1_{q,q} \end{bmatrix} \quad (3.35)$$

$$= \sum_{i=1}^q \sum_{j=1}^q R_{i,j} W1_{i,j}. \quad (3.36)$$

This operation functions similar to calculating the probability-weighted average of the state $R_{i,j}$:

$$E(\mathbf{R}) = \sum_{i=1}^Q \sum_{j=1}^Q R_{i,j} \rho(R_{i,j}) \quad (3.37)$$

where $E(\mathbf{R})$ is the expected value of the matrix \mathbf{R} in the Poincaré plot on the training data. Therefore, the mask processing between \mathbf{R} and $\mathbf{W1}$ aims to approximate the expected value of \mathbf{R} :

$$\mathbf{R} * \mathbf{W1} \approx E(\mathbf{R}). \quad (3.38)$$

After filtered by the mask $\mathbf{W1}$, certain features from the matrix \mathbf{R} would be extracted based on the distribution of $RR(n)$. In other words, a sequence of $RR(n)$ from the AF sinus rhythm would show more similarity to typical AF sinus rhythm. Its states in the matrix \mathbf{R} that are likely to be visited by an AF signal would be detected by the mask $\mathbf{W1}$ and lead to a higher expected value at last, while a new data of normal sinus rhythm has fewer states in general that can be detected by the filter, and would result in a lower expected value.

3.3.3. Position of bins in the Poincaré plot

In addition to $\mathbf{W1}$, another mask $\mathbf{W2}$ is also implemented to help with the classification. Let $\mathbf{W2}$ be a $Q \times Q$ matrix. The elements of $\mathbf{W2}$ consists of the Euclidean distance of each point to the center. Therefore, elements on the edge have a higher value than the elements around the center. The visualization of the mask $\mathbf{W2}$ can be seen in Figure 3.9.

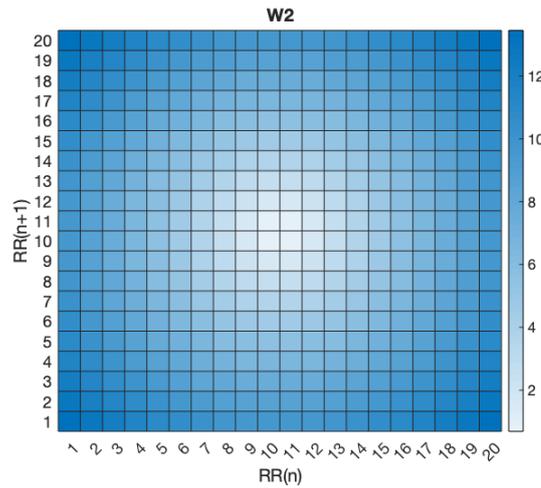


Figure 3.9: Mask $\mathbf{W2}$.

Again, the element-wise multiplication and summation between the matrix \mathbf{R} and the mask $\mathbf{W2}$ (denoted as $\mathbf{R} * \mathbf{W2}$) is given as:

$$\mathbf{R} * \mathbf{W2} = \begin{bmatrix} R_{1,1} & \cdots & R_{1,Q} \\ \vdots & \ddots & \vdots \\ R_{Q,1} & \cdots & R_{Q,Q} \end{bmatrix} * \begin{bmatrix} W2_{1,1} & \cdots & W2_{1,Q} \\ \vdots & \ddots & \vdots \\ W2_{Q,1} & \cdots & W2_{Q,Q} \end{bmatrix} \quad (3.39)$$

$$= \sum_{i=1}^Q \sum_{j=1}^Q R_{i,j} W2_{i,j}. \quad (3.40)$$

Because of the normalization of the RR intervals, the average value of the normalized RR intervals is one. It corresponds to the $(RR_{\frac{Q}{2}}, RR_{\frac{Q}{2}})$ combination in the Poincaré plot. For Normal rhythm, most combinations are around $(RR_{\frac{Q}{2}}, RR_{\frac{Q}{2}})$, while for AF rhythm, the situation becomes complicated and the combinations could have more outliers. This means for both the rhythms, the element $R_{\frac{Q}{2}, \frac{Q}{2}}$ in the

matrix \mathbf{R} is the state most likely to be visited by the system. Other than this, for the Normal rhythm, states around $R_{\frac{Q}{2}, \frac{Q}{2}}$ are likely to be visited by the system and thus are set to be one in the matrix \mathbf{R} . However, for the AF rhythm, the matrix \mathbf{R} has more states that are far from the center and still likely to be visited by the system. By giving these outliers higher weights, the mask $\mathbf{W2}$ aims to detect the non periodic characteristics of the AF ECG recordings. Therefore, in the mask $\mathbf{W2}$, the weight is chosen as the Euclidean distance between each state $R_{i,j}$ and the center of the Poincaré plot.

After element-wise multiplying \mathbf{R} with matrix $\mathbf{W2}$ (denoted as $\mathbf{R} \circ \mathbf{W2}$), an element in the matrix \mathbf{R} that is far from the center will have a higher weight, while an element close to the center will have a lower weight. This also complies with the intention to give a high weight for AF signal and a low weight for Normal signal, because an AF signal shows a higher dispersion than Normal signal and thus has more points far from the center. The example are shown in Figure 3.10, where the matrix $\mathbf{R} \circ \mathbf{W2}$ of a Normal signal has entries ranging from 0 to 2, and the matrix $\mathbf{R} \circ \mathbf{W2}$ of an AF signal has entries ranging from 0 to 7. Finally, the summation of all the elements after element-wise multiplication leads to a higher value for AF rhythm, due to the outliers with higher weights; and a lower value for Normal rhythm, due to the states around the center with lower weights.

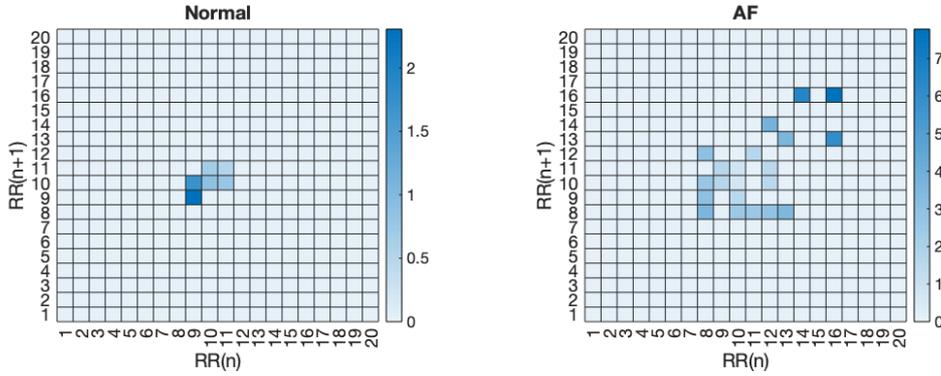


Figure 3.10: Results of element-wise multiplication of $\mathbf{R} \circ \mathbf{W2}$. An example of Normal ECG recordings is shown on the left. An example of AF ECG recordings is shown on the right.

3.3.4. Combination of distribution and position

Given that the masks $\mathbf{W1}$ and $\mathbf{W2}$ are calculated in the previous subsections, both of them can be element-wise multiplied and summarized with \mathbf{R} together as one feature to reduce the dimension of the feature space. Let F_7 be the new feature this paper proposes - weighted area of bins:

$$F_7 = \mathbf{R} * (\mathbf{W1} \circ \mathbf{W2}) \quad (3.41)$$

$$= \begin{bmatrix} R_{1,1} & \cdots & R_{1,Q} \\ \vdots & \ddots & \vdots \\ R_{Q,1} & \cdots & R_{Q,Q} \end{bmatrix} * \begin{bmatrix} W_{1,1}W_{2,1,1} & \cdots & W_{1,Q}W_{2,1,Q} \\ \vdots & \ddots & \vdots \\ W_{1,Q,1}W_{2,Q,1} & \cdots & W_{1,Q,Q}W_{2,Q,Q} \end{bmatrix} \quad (3.42)$$

$$= \sum_{i=1}^Q \sum_{j=1}^Q R_{i,j} W_{1,i,j} W_{2,i,j} \quad (3.43)$$

where \circ is element-wise multiplication for matrix, and $*$ contains element-wise multiplication and summation. Generally speaking, feature F_7 combines the weights from the masks $\mathbf{W1}$ and $\mathbf{W2}$. In this way, while approximating the expected value of the matrix \mathbf{R} , the feature F_7 also gives higher weights to the outliers from the center of the Poincaré plot, which are likely to occur during the AF rhythm, thus leading to a larger difference for the feature value between the two classes.

Compared with SD1 and SD2, F_7 explores the nonlinear dynamics of the Poincaré plot and the resolution of the plot can be tuned by changing the level of quantization Q ; Compared with F_4 and F_5 , F_7 aims to consider the distribution and location information of the states visited by the system, which makes sure the irregularity pattern of AF sinus rhythm is better exploited.

3.4. Autocorrelation function

In the previous section, the Poincaré plot for RR intervals is analyzed. In this section, the thesis proposes that the Poincaré plot can also be calculated from the autocorrelation function.

The autocorrelation measures the correlation of a signal with itself across time. It is defined as:

$$ACF_{XX}(m) = E[X_{n+m}X_n^*] \quad (3.44)$$

In this thesis, the autocorrelation function \widehat{ACF}_{XX} of a vector X with length of N is estimated as:

$$\widehat{ACF}_{XX}(m) = \frac{1}{\sum_{n=0}^{N-1} X_n X_n^*} \sum_{n=0}^{N-m-1} X_{n+m} X_n^* \quad (3.45)$$

where $m = 0, 1, 2, \dots, N - 1$, the asterisk denotes complex conjugation, and E is the expected value operator [11]. Because the autocorrelation function is symmetric about the origin, only $m \geq 0$ part is taken into account.

At first, the ECG recordings are segmented into 5 second windows. The step of the sliding window is 50% of the size of the window, 2.5 seconds. Then, for each window, a 2-order Butterworth high-pass filter with cut-off frequency of 0.5 Hz is applied to remove the baseline wander, which is computationally cheap and almost as accurate as other baseline wander removal methods [27]. The effect of baseline wander removal by this filter can be seen in Figure 3.11.

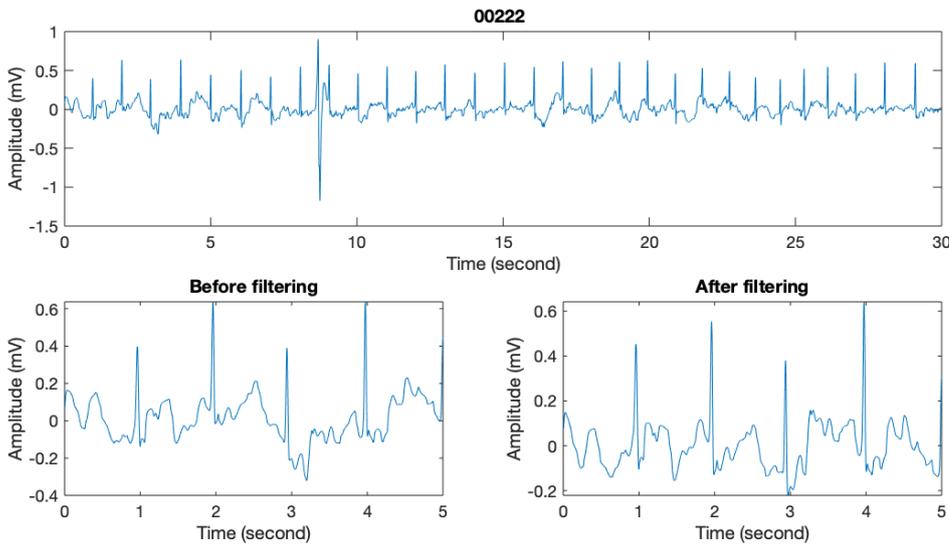


Figure 3.11: An example of Normal sinus rhythm showing the original signal (bottom left) and the filtered signal (bottom right).

3.4.1. Autocorrelation of ECG

In Figure 3.12, a filtered segment of Normal and AF sinus rhythm is shown on the left. The autocorrelation of the signal is then calculated in the middle. It can be clearly seen that compared with AF ECG signals, the peaks in the autocorrelation function of Normal rhythm are higher and more periodic. This is because of the irregular pattern of the AF rhythm which causes the R peaks do not coincide with each other when shifting. To deal with the situation where peaks are too close to each other or too low and therefore do not show cyclic characteristics, many combinations of defining the peak height and the peak distance are visually inspected, and only peaks that are higher than 0.1 and at least 100 lags away from each other are finally selected (implying that the heart rate is assumed to be lower than 180 bpm) and marked as red circles in the plot.

After all the peaks of the autocorrelation function from the sliding window for one recording are detected, a sequence of peak intervals, denoted as $PP(n)$, are obtained. The sequence can be used to generate the Poincaré plot as it was done for $RR(n)$. This is shown in Figure 3.12 on the right. Again, the Poincaré plot of AF rhythm shows a higher dispersion than that of Normal rhythm. This method is

based on calculating the repeating pattern of the ECG signal, instead of simply counting the heart rate. Therefore, it shows more information hidden in the morphology feature.

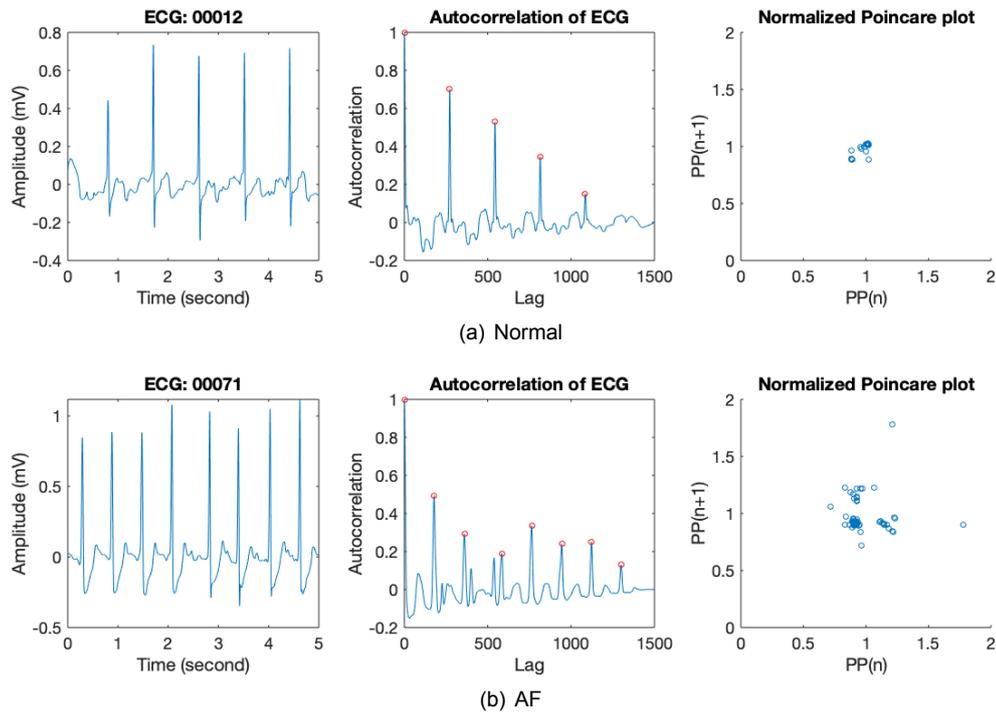


Figure 3.12: Examples of Normal and AF rhythm recordings showing how the peaks in autocorrelation function are defined and how the Poincaré plot is generated.

3.4.2. Autocorrelation of the prediction error

In this section, the autocorrelation of the prediction error is calculated instead of the autocorrelation of the ECG. The autocorrelation of the prediction error provides another perspective to model an ECG signal. The concept of prediction error in this scenario comes from linear predictive coding (LPC), which appears to be powerful for many applications including speech signal processing and ECG signal processing. LPC predicts the current sample of a signal based on a linear combination of previous samples of the signal [35], that is,

$$\tilde{x}(n) = \sum_{i=1}^q b_i x(n-i) \quad (3.46)$$

where $\tilde{x}(n)$ is the predicted value, $x(n-i)$ is the previous observation, b_i is the LPC coefficient, and q is the order of LPC. The prediction error is then the difference between the predicted value $\tilde{x}(n)$ and the measured value $x(n)$:

$$e(n) = \tilde{x}(n) - x(n) \quad (3.47)$$

and it is minimized in the least squares sense to get the LPC coefficients. Given the coefficients b_i , one can construct the prediction error filter $B(z)$, which can be used to calculate the prediction error by filtering the ECG signal with $B(z)$, that is,

$$B(z) = 1 - \sum_{i=1}^q b_i z^{-i}. \quad (3.48)$$

In Figure 3.13, a filtered segment of the Normal and AF sinus rhythm after baseline wander removal is again shown in the first column. This is then filtered by the prediction error filter $B(z)$, where $q = 10$. The prediction error is shown in the second column. It can be seen that in the prediction error, short-term correlation within q samples is partly reduced, and the major variation occurs at the same place of the QRS complex. This can reduce the influence of P wave and T wave to some extent.

In the third column, the autocorrelation of the prediction error is calculated. Just like the autocorrelation of the ECG, the autocorrelation of the prediction error also has higher and more periodic peaks for Normal rhythm, and more irregular and uneven peaks for AF signal. In general, the peaks in the autocorrelation of the prediction error are not as obvious as those in the autocorrelation of the ECG, which is probably due to the high variation of the prediction error. Therefore, the peaks in this case are defined when higher than 0.05 and having a minimum distance of 100 lags (marked in red circles).

After all the peaks of the autocorrelation function of prediction error from the sliding window for one recording are detected, a sequence of peak intervals, denoted as $PP1(n)$, are obtained. $PP1(n + 1)$ is plotted against $PP1(n)$ to get the Poincaré plot, which is shown in the fourth column in Figure 3.13. The plot in this case also shows that the AF signal has a higher dispersion than the Normal signal, and therefore it is again assumed to be a potential feature to detect AF from Normal.

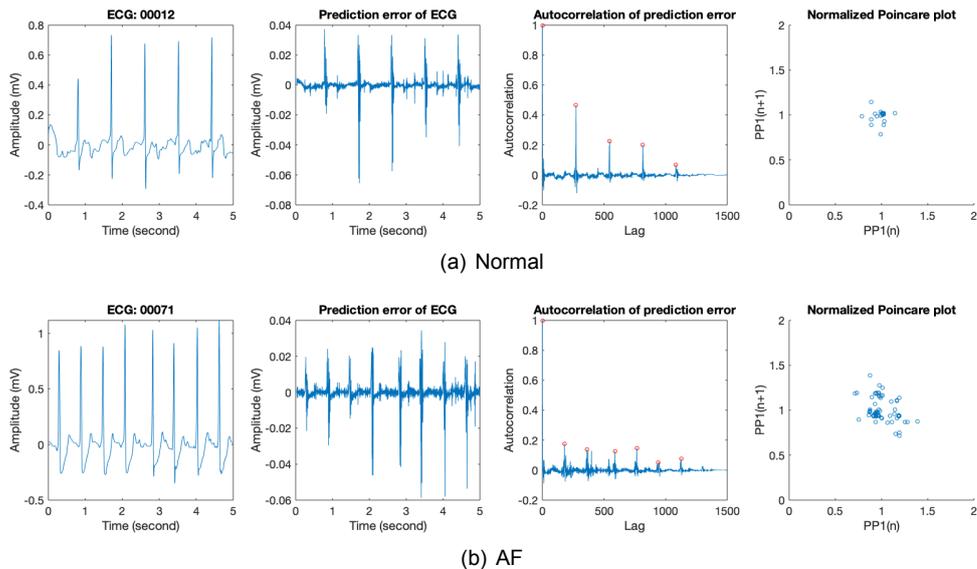
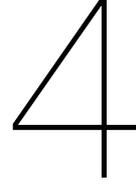


Figure 3.13: Examples of Normal and AF rhythm recordings showing how the peaks in autocorrelation function of prediction error are defined and how the Poincaré plot is generated.



Feature selection and classification

In this chapter, features generated from the Chapter 3 are selected and used to classify Normal and AF signals. In Section 4.1, the minimum redundancy maximum relevance algorithm is applied to rank the extracted features based on their importance. In Section 4.2 and 4.3, support vector machines and random forests are introduced to perform the classification task.

4.1. Minimum redundancy maximum relevance algorithm

After features are extracted, part of the features are often found to be irrelevant or redundant, which can reduce the performance of the classifier. In this thesis we use the minimum redundancy maximum relevance (mrmr) algorithm to select features. It aims to select a subset T of all the features with a high correlation with the class and a low correlation with each other. The algorithm uses mutual information I of variables to represent redundancy U_x and relevance V_x of a feature x [44]:

$$I(X, Y) = \sum_{x, y \in T} \rho(x, y) \log \left(\frac{\rho(x, y)}{\rho(x)\rho(y)} \right) \quad (4.1)$$

$$U_x = \frac{1}{|T|} \sum_{x, z \in T} I(X, Z) \quad (4.2)$$

$$V_x = I(X, Y) \quad (4.3)$$

where $\rho(x, y)$ is the joint probability density, $\rho(x)$ and $\rho(y)$ are the marginal probability density functions, Z is the class label, X and Y are features, and $|T|$ is the number of features in the set T . Features are selected one-by-one using a greedy search to maximize the criterion of feature importance, mutual information quotient (MIQ):

$$MIQ_x = \frac{V_x}{U_x} \quad (4.4)$$

The mrmr feature selection process is as follows [10]:

1. The feature with the highest relevance is selected into the feature set T
2. The feature with the highest relevance and zero redundancy in the complement of T , denoted as T^c , is selected into T
3. Repeat step 2 until no feature in T^c has zero redundancy
4. The feature with the highest MIQ value in T^c is selected into T
5. Repeat step 4 until no feature in T^c has non-zero relevance
6. Features with zero relevance are added into T in random order

The ranking of all the features for the whole dataset using mrmr algorithm is shown in the following figure.

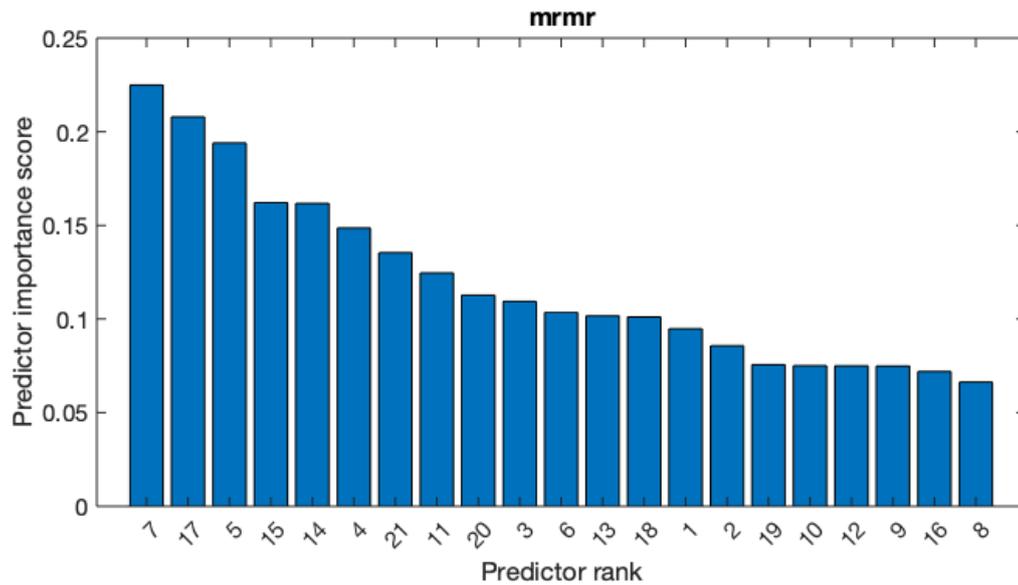


Figure 4.1: Results of mrmr algorithm.

The indices of the features in Figure 4.1 are given in the table below and the corresponding rankings are shown in the bracket.

Table 4.1: index of features

Features	RR intervals	Peak intervals in acf of ECG	Peak intervals in acf of prediction error
SD1 (F_1)	1 (14)	8 (21)	15 (4)
SD2 (F_2)	2 (15)	9 (19)	16 (20)
SD1/SD2 (F_3)	3 (10)	10 (17)	17 (2)
Mean stepping increment (F_4)	4 (6)	11 (8)	18 (13)
Complex correlation measure (F_5)	5 (3)	12 (18)	19 (16)
Unweighted area of bins (F_6)	6 (11)	13 (12)	20 (9)
Weighted area of bins (F_7)	7 (1)	14 (5)	21 (7)

It can be seen in the table 4.1 that among the seven kinds of features from RR intervals, peak intervals in acf of ECG, and those of prediction error, four of them have the highest ranking for RR intervals, and the other three of them have the highest ranking for peak intervals in acf of prediction error. While peak intervals in acf of ECG do not have the highest ranking feature for any of the seven kinds of features, some of them rank relatively high among the whole 21 features. For example, F_7 and F_4 rank the fifth and eighth, respectively.

When analyzing the features for RR intervals only, traditional features (F_1 , F_2 and F_3) have a low score, while the features that explore the temporal structure (F_4 and F_5) and distribution (F_7) score higher. When looking at only the features for peak intervals in the acf of the ECG, F_7 and F_4 seem more important than others. As for the features for peak intervals in acf of prediction error, traditional features

(F_1 and F_3) and new features (F_6 and F_7) rank in the first half of all the features, and the temporal related features, on the other hand, do not rank high. In all of these three domains, F_7 shows a promising score, which means the distribution of data in the Poincaré plot and the relative position of states visited by the system are helpful with classification of AF signals from Normal signals.

4.2. Support vector machines

The Support vector machines (SVMs) algorithm is a widely used classifier for both linear and nonlinear (non-)separable data in the pattern recognition field. The goal of SVM is to create a hyperplane that makes sure that the distance from the nearest point of every class to the hyperplane is maximized [16, 39]. The hyperplane is defined as:

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = 0 \quad (4.5)$$

where \mathbf{w} determines the direction of the hyperplane, w_0 determines its exact position and \mathbf{x}_i ($i = 1, \dots, N$) is the feature vector of the dataset X , which belongs to the classes w_1 or w_2 . The distance of a point to the hyperplane is given by:

$$z = \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|}. \quad (4.6)$$

By scaling \mathbf{w} and w_0 , the value of $g(\mathbf{x})$ is equal to one at the nearest points for w_1 , and minus one at the nearest points for w_2 . For linearly separable data, the goal can be expressed as maximizing the margin between the two classes:

$$\frac{1}{\|\mathbf{w}\|} + \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (4.7)$$

with the constraints that

$$\mathbf{w}^T \mathbf{x} + w_0 \geq 1, \quad \forall \mathbf{x} \in w_1 \quad (4.8)$$

$$\mathbf{w}^T \mathbf{x} + w_0 \leq -1, \quad \forall \mathbf{x} \in w_2. \quad (4.9)$$

By representing $y_i = 1$ for $\mathbf{x}_i \in w_1$ and $y_i = -1$ for $\mathbf{x}_i \in w_2$, the objective function can be stated as:

$$\min \quad J(\mathbf{w}, w_0) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (4.10)$$

$$\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1, \quad i = 1, \dots, N. \quad (4.11)$$

Let $L(\mathbf{w}, w_0, \boldsymbol{\lambda})$ be the Lagrangian function where $\boldsymbol{\lambda}$ is the vector of the Lagrange multipliers:

$$L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \lambda_i [y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1]. \quad (4.12)$$

According to the Karush–Kuhn–Tucker (KKT) conditions for optimality, the following constraints have to be satisfied:

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \mathbf{0} \quad (4.13)$$

$$\frac{\partial}{\partial w_0} L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = 0 \quad (4.14)$$

$$\lambda_i \geq 0, \quad i = 1, \dots, N \quad (4.15)$$

$$\lambda_i [y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1] = 0, \quad i = 1, \dots, N. \quad (4.16)$$

Substituting equation 4.12 into equation 4.13 and 4.14 results in

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i \quad (4.17)$$

$$\sum_{i=1}^N \lambda_i y_i = 0. \quad (4.18)$$

The problem can then be constructed as its equivalent Lagrangian duality form:

$$\max_{\lambda} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \quad (4.19)$$

$$\text{subject to } \sum_{i=1}^N \lambda_i y_i = 0 \quad (4.20)$$

$$\lambda \geq 0. \quad (4.21)$$

Once λ is calculated, \mathbf{w} can be obtained from equation 4.17, and w_0 can be obtained from equation 4.16.

When the data can not be separated linearly, slack variables ξ_i are introduced to penalize the outliers that are misclassified. The objective function becomes:

$$\min J(\mathbf{w}, w_0, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (4.22)$$

$$\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 - \xi_i, \quad i = 1, \dots, N \quad (4.23)$$

$$\xi_i \geq 0, \quad i = 1, \dots, N \quad (4.24)$$

where C is a positive constant to balance between the width of the margin and the number of misclassified points. The corresponding dual representation is formulated as:

$$\max_{\lambda} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \quad (4.25)$$

$$\text{subject to } \sum_{i=1}^N \lambda_i y_i = 0 \quad (4.26)$$

$$0 \leq \lambda_i \leq C, \quad i = 1, \dots, N \quad (4.27)$$

When the number of data in each class is unbalanced, which is the case in this thesis, the penalty parameter C for different classes is multiplied with a weight that is reversely proportional to the size of that class. The idea is to penalize heavily the outliers from the class with less data, so that the classifier can reduce its bias towards the class with more data.

For the nonlinear case, a kernel function is often used to project the input feature space into a higher dimension, where the classes are more easily separated by a hyperplane. This is achieved by expressing the inner product of the vectors in the higher dimension space as a function of the inner product of the corresponding vectors in the original feature space [39]. Commonly used kernel functions from the pattern recognition field include:

$$\text{Polynomials : } K(x, z) = (x^T z + 1)^q, \quad q > 0 \quad (4.28)$$

$$\text{Radial Basis Functions : } K(x, z) = \exp(-\gamma \|x - z\|^2) \quad (4.29)$$

$$\text{Sigmoid : } K(x, z) = \tanh(\beta x^T z + \gamma) \quad (4.30)$$

After a certain kernel function is selected, the dual problem becomes:

$$\max_{\lambda} \left(\sum_i \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (4.31)$$

$$\text{subject to } 0 \leq \lambda_i \leq C, \quad i = 1, \dots, N \quad (4.32)$$

$$\sum_i \lambda_i y_i = 0. \quad (4.33)$$

4.2.1. Searching for optimal parameters

In this thesis, Radial Basis Function (RBF) is selected as the kernel function of SVM. This is because the RBF kernel can nonlinearly map the dataset into a higher dimensional space, so that the SVM classifier works better with the nonlinear data. Another reason is the number of parameters which affects the complexity of the classifier. For RBF kernel, there is only one parameter γ , while polynomial kernel has much more parameters. Furthermore, sigmoid kernel is not chosen because it behaves like RBF kernel for certain parameters, and it is not valid in some case [25]. Therefore, when RBF kernel is used, the penalty parameter C and RBF kernel parameter γ need to be tuned during the training procedure.

In order to search for the optimal parameters C and γ , a grid search using 10-fold cross-validation is used in this thesis. For a certain parameter pair (C, γ) value, a 10-fold cross-validation is implemented to get an average performance. This is illustrated in Figure 4.2. The whole training set is first divided into ten subsets with the same size. The first subset is selected as validation set, while the other nine subset is used as training set. The SVM classifier is trained on the training set and then evaluated on the validation set. The accuracy of the classifier is recorded as Acc_1 and the corresponding model is discarded. Next, the second subset is used as the validation set, while the other 9 subset is used for training the SVM classifier. This is repeated ten times until every single subset has been chosen once as the validation set.

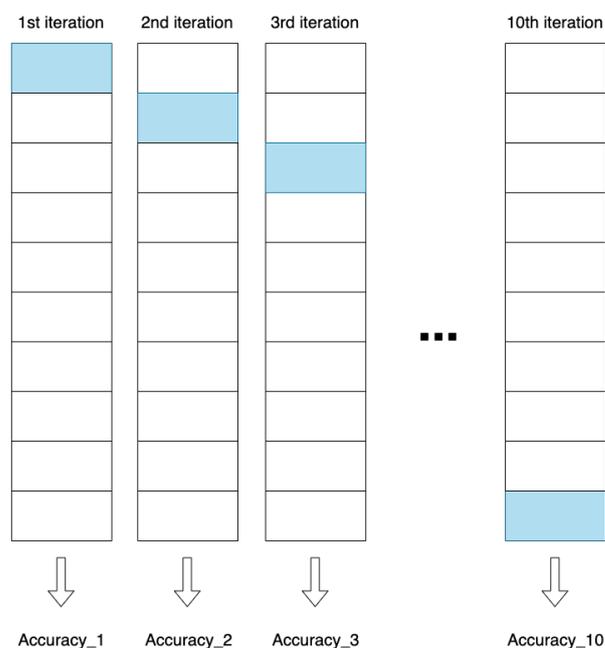


Figure 4.2: Diagram of 10-fold cross validation. Training set is denoted as white. Validation set is denoted as blue.

When all ten accuracies are obtained, the cross-validation accuracy Acc is calculated as the average of all the accuracies:

$$Acc = \frac{1}{10} \sum_{i=1}^{10} Acc_i \quad (4.34)$$

This is the performance for a certain parameter pair. According to [25], it is reasonable to search for the parameters exponentially. Therefore, the range for these two parameters in this thesis is defined as $C \in [2^1, 2^9]$ and $\gamma \in [2^{-7}, 2^1]$ with the step size being 2^2 , as shown in Figure 4.3.

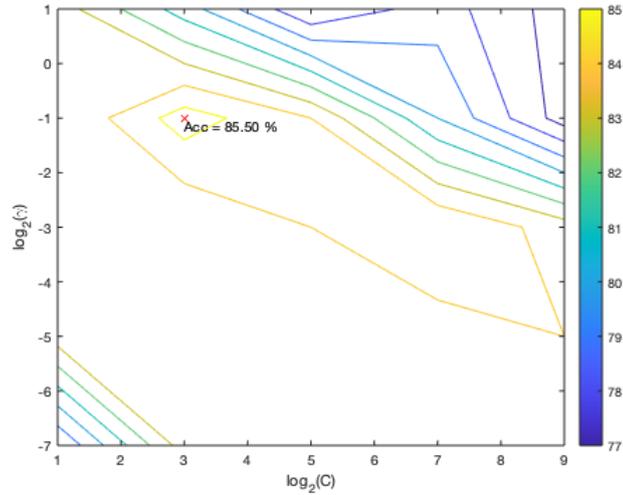


Figure 4.3: Grid search for the optimal C and γ .

After the cross-validation accuracy for every (C, γ) pair is calculated, the optimal (C, γ) is selected from the SVM model with the highest accuracy (marked as a red cross in the figure). They are then used to model the final SVM classifier on the entire training set.

4.3. Random forest

Besides the SVM classifier, this thesis report also uses random forest (RF) as comparison to see which classifier can reach a better result. Random forest is an ensemble learning method consisting of multiple decision trees as basic classifiers. This section starts with the concept of decision trees, followed by the algorithm of random forest.

4.3.1. Decision tree

A decision tree is a tree-like non-linear supervised classifier consisting of a sequence of decision stages until a final class is accepted, as illustrated in Figure 4.4. Each block (denoted as a decision node) in the figure represents a question asking if a certain feature is larger or smaller than a value. The answers 'yes' and 'no' split a node into two different sub-nodes with no overlapping. The first node (denoted as the root node) of the decision tree is associated with the whole training set. The node at the end of a branch (denoted as a leaf node) contains data from the same class.

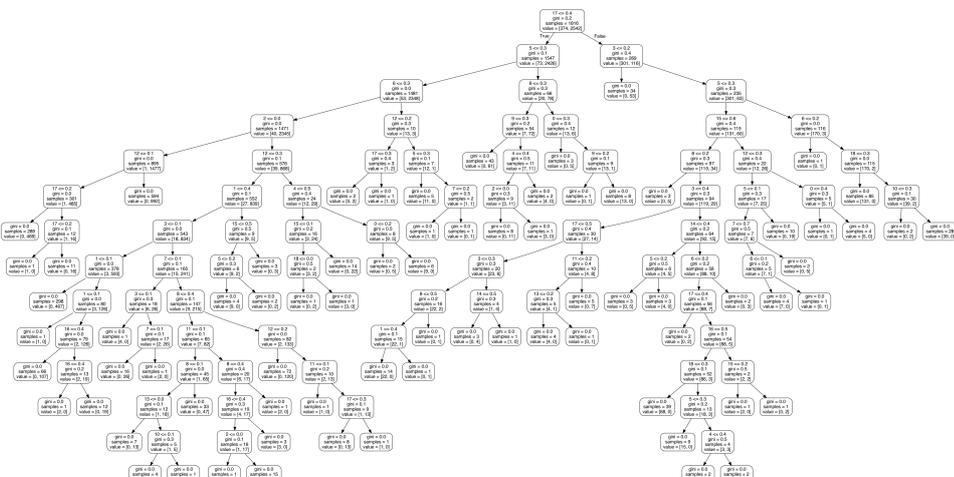


Figure 4.4: Diagram of a decision tree.

In general, the decision trees aim to divide the data so that after the splitting, data coming from different classes are divided into different child nodes as much as possible. In other words, the purity of the data in the child nodes becomes higher than in the parent nodes. The commonly used splitting criterion at each decision node includes 'ID3', 'C4.5', and 'CART'. In this thesis, CART (Classification and Regression Tree) is applied to measure the purity of the data set. It uses Gini index to create binary splits. Lower value of Gini means higher homogeneity. Suppose the proportion of the k th class in the data set D is P_k ($k = 1, 2, \dots, |K|$), Gini index of the data set D is then calculated as:

$$Gini(D) = \sum_{k=1}^{|K|} \sum_{k' \neq k} P_k P_{k'} \quad (4.35)$$

$$= 1 - \sum_{k=1}^{|K|} P_k^2 \quad (4.36)$$

It reflects the probability of selecting two different samples randomly from the data set D . Suppose a feature a has V possible values $\{a^1, a^2, \dots, a^V\}$, if a is chosen to split the data set D , there will be V possible splitting point. All the samples with the feature $a = a^v$ are denoted as D^v . Gini index of a feature a is defined as:

$$Gini(D, a) = \sum_{v=1}^V \frac{|D^v|}{D} Gini(D^v) \quad (4.37)$$

where $\frac{|D^v|}{D}$ is the weight for the splitting point a^v . Therefore, in the feature set A , the optimal splitting feature a^* is the one leads to the smallest Gini index after splitting:

$$a^* = \underset{a \in A}{\operatorname{argmin}} Gini(D, a). \quad (4.38)$$

The algorithm for growing a decision tree is summarized as follows:

1. The algorithm begins with the root node which contains the whole training set.
2. For each node, Gini index for each feature and each value is calculated. The feature and associated splitting value leading to the smallest Gini index are chosen and two descendant nodes are generated.
3. Declare a node as a leaf if all the data belong to the same class or the cardinality of the data is small enough or the depth of the branch reaches a certain value. Otherwise, repeat step 2.

The algorithm of decision trees is straightforward and easy to interpret. It can handle both numerical and categorical data. It performs well on large datasets [8]. However, decision trees are prone to overfitting. This issue can be mitigated by setting a max depth of the tree. But, this could increase error rate due to bias. Another solution to lessen the overfitting issue is random forest.

4.3.2. Random forest

Random forests combine the idea of bagging together with random feature selection [13]. Random forests are based on the growth of various different decision trees. The samples used to grow a decision tree are selected randomly from the whole dataset with replacement. This makes sure each decision tree has different input data. As decision trees are sensitive to the training set, even small changes lead to different tree structures. Another randomness for random forest comes from the feature selection. For each node in a decision tree, the optimal feature is chosen from a randomly selected subset of the whole feature set. Without the random feature selection, certain predictive features would be chosen in many base trees, causing these trees in the forest to become highly correlated.

These two kinds of randomness make a random forest combat error due to variance. A number of relatively uncorrelated models work together can outperform any single one of them. While one tree might grow too deep and tend to have a low bias and a high variance, various trees together average the multiple results and reduce the variance, with a small increase of the bias. In practice the variance

reduction is often significant. When having a new data, the output of a random forest is the one with the most votes of the decision trees. Therefore, a random forest is less likely to overfit than a decision tree.

In this thesis, parameters of the random forest algorithm are set to be the default value of the 'scikit-learn' machine learning library, since searching for the optimal parameters does not witness a noticeable improvement. The number of trees in the forest is set to be 100, which makes sure the performance generalize well to different unknown data, yet do not increase the complexity of the algorithm too much. The number of features to consider when looking for the best split is set to be the square root of the number of the total features. This makes each tree grow differently from each other, and still includes the important features for the majority of the trees at the same time. The weight for each class is set to be one. This already generates a satisfying result, and giving more weight to the class with less data does not lead to better results.

5

Results

In this chapter, performance of the SVM and RF classifiers using the Poincaré plot based features is presented. In Section 5.1, the evaluation metrics used in this thesis are introduced. In Section 5.2, the choice of the quantization level Q in the Poincaré plot is discussed. In Section 5.3, the distribution of all the features including the data quartiles and medians are compared in the box plot. In Section 5.4, the classification results using RR intervals and/or peak intervals in autocorrelation are shown. In Section 5.5, feature selection is implemented to see if the classification results improve. In Section 5.6, m-lagged analysis for the feature F_7 is discussed.

5.1. Evaluation metrics

Table 5.1: Confusion matrix

	Predicted Positive (class 1)	Predicted Negative (class 2)
Actual Positive (class 1)	True positive (TP)	False negative (FN)
Actual Negative (class 2)	False positive (FP)	True negative (TN)

Table 5.1 shows a direct way to present the results of a classification model - Confusion matrix, which consists of true positive, true negative, false negative, and false positive respectively. Each row shows the actual labels in the dataset, while each column shows the predicted label generated by the classifier. So, for example, TP includes all the data that belong to class 1 and are predicted as class 1 by the classifier.

Four commonly used evaluation metrics that can be calculated from a confusion matrix are accuracy, precision, recall, and F1 score. Accuracy is the proportion of all the correctly classified data in the whole dataset:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}. \quad (5.1)$$

Precision is the proportion of the true positive in the predicted positive:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (5.2)$$

Recall is the proportion of the true positive in the actual positive:

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (5.3)$$

F1 score is a combination of Precision and Recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (5.4)$$

Calculating accuracy is an intuitive way to measure how accurate a classification model is, but sometimes it can be misleading, especially when the class labels are imbalanced. When a model predicts correctly the majority of the class with a large size but fails in predicting the class with a small size, accuracy can still be high, but this does not mean it is a good model. Precision shows that out of all the predicted positive samples, how many of them are indeed positive. This could be useful when the cost of false positives is high. Recall calculates how many actual positives are indeed predicted as positive by the model. This is useful when the cost is high for false negative. F1 score, on the other hand, balances between Precision and Recall, and is more reliable when the data class is uneven [37].

5.2. Choice of quantization level Q

From section 3.2.3 to 3.3, features F_6 and F_7 quantized the Poincaré plot into Q levels in order to extract the distribution information and the position information of a certain (RR_i, RR_j) combination that occurred in the plot. However, exact value of Q has not been specified until the classifiers are introduced.

In this section, different values of Q are chosen and the SVM and RF classifiers are used with all the 21 features. To make sure the ratio of the two classes is consistent in both the training set and the test set, 50% of each class in the dataset is selected randomly to model the SVM classifier, and the other 50% is used to evaluate the performance of the model. The optimal value of Q will be chosen as the one that can lead to the highest F1 score in the test set. Results can be seen in Table 5.2.

Table 5.2: Choice of Q

Value of Q	length of the bin	SVM Accuracy	SVM F1 score	RF Accuracy	RF F1 score
5	0.4	94.47%	0.79	95.41%	0.82
10	0.2	94.10%	0.78	96.40%	0.86
20	0.1	94.65%	0.80	96.56%	0.87
25	0.08	94.96%	0.81	96.69%	0.87
40	0.05	95.09%	0.82	96.80%	0.87
50	0.04	95.06%	0.82	96.76%	0.88
100	0.02	95.47%	0.83	96.85%	0.88
200	0.01	92.83%	0.68	95.11%	0.78
400	0.005	88.78%	0.31	90.89%	0.48
800	0.0025	86.86%	0.06	87.82%	0.15

It can be seen from the table that as the value of Q increases from 5 to 100, the F1 score also rises from 0.88 to 0.90. When the value of Q further increases, the F1 score drops, and the running time of the algorithm increases noticeably. The possible reason of the low F1 score with a high value of Q could be due to the deficiency in the size of the training set. The training set only contains around three thousand of ECG recordings, of which less than four hundred recordings belong to AF rhythm. When the value of Q is too high, many states might not even be visited by the system, leading to an unsatisfying results of the F1 score. Therefore, in the following sections, the value of Q is set to be 100.

5.3. Comparison of the distribution of the features

In this section, the distribution of all the features are compared with each other, in respect of the RR intervals, the peak intervals in the autocorrelation of ECG, and the peak intervals in the autocorrelation of the prediction error. The most extreme data, the quartiles and the medians are presented in the box plot. The means and the standard deviations are presented in the table. All the features are again presented in Table 5.3.

Table 5.3: Brief explanation of the features

Annotation	Features	Explanation
F_1	SD1	Standard deviation along the identity line in the Poincaré plot
F_2	SD2	Standard deviation along the direction perpendicular to the identity line
F_3	SD1/SD2	Ratio of SD1 and SD2
F_4	Mean stepping increment	Mean distance between the consecutive points in the Poincaré plot
F_5	Complex correlation measure (CCM)	Mean area of the triangles composed of the consecutive three points in the Poincaré plot
F_6	Unweighted area of bins	Count of non-zero bins in the Poincaré plot
F_7	Weighted area of bins	Giving weights to the bins in feature F_6

5.3.1. RR intervals

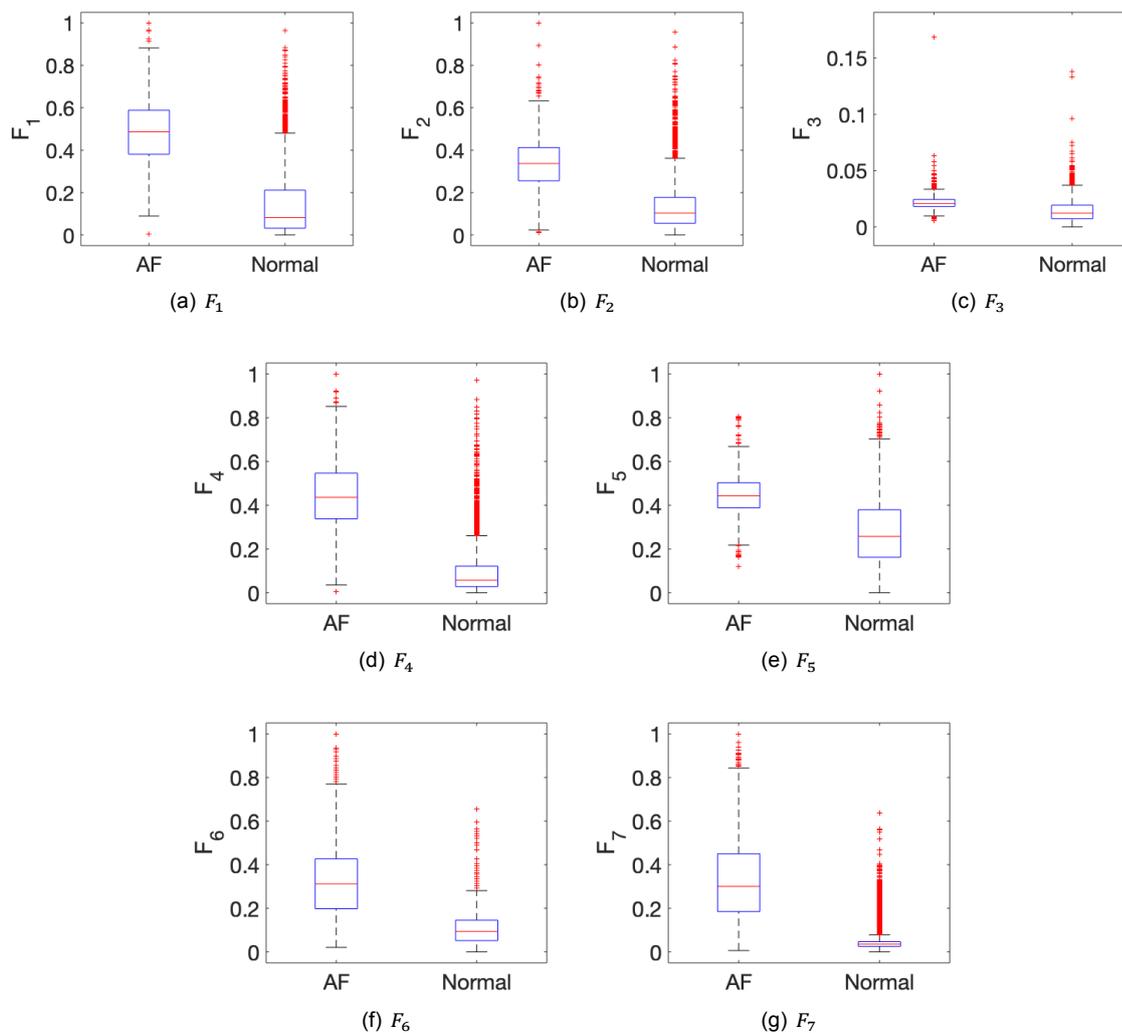


Figure 5.1: Comparison of features.

Figure 5.1 illustrates the comparison of distribution of features F_1 to F_7 using the box plot, where the bottom and top edges of the box indicate the 25th and 75th percentiles of the feature values, and the central mark indicates the median of the feature value. The whiskers indicate the most extreme data points that are not considered as outliers, and the outliers are plotted using the red '+' symbol [4].

It can be seen from Figure 5.1 that except for feature F_3 , there is no overlap between the two classes, although the boxes for feature F_5 are very close to each other. For features F_1 , F_2 , and F_6 , the upper whisker of the Normal rhythm is overlapping with the box of the AF rhythm. The most separable features are F_4 and F_7 , where the upper whisker of the Normal rhythm is obviously not overlapping with the box of the AF rhythm.

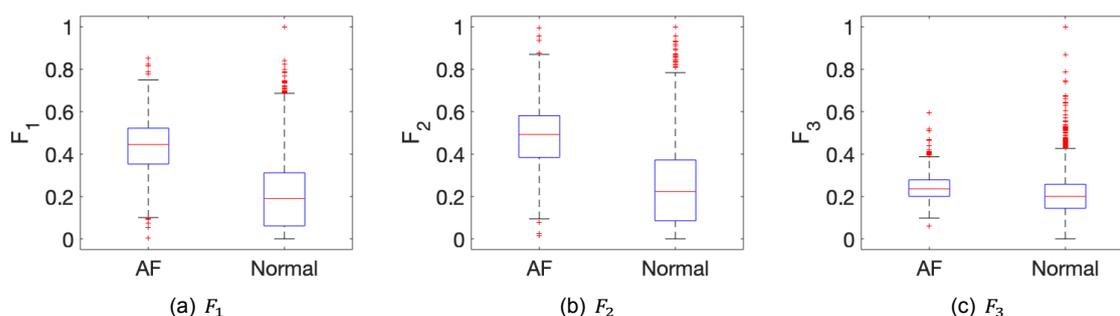
The mean value, standard deviation value, and p-value from the Mann–Whitney U test of each feature are shown in Table 5.4. Among these seven features, the mean value of feature F_3 for the two classes is the closest. The mean value of other features for the AF class is around twice or three times of that for the Normal class, while the mean value of the feature F_7 for the AF class is more than five times of that for the Normal class. All the p-values are much smaller than 0.001, meaning there is significant difference between the two classes for these features.

Table 5.4: Mean \pm standard deviation of all features

Feature	Normal	AF	p-value
F_1	0.14 ± 0.15	0.48 ± 0.15	2.44e-165
F_2	0.14 ± 0.12	0.35 ± 0.13	1.88e-135
F_3	0.01 ± 0.01	0.02 ± 0.01	1.31e-62
F_4	0.10 ± 0.11	0.44 ± 0.15	5.25e-184
F_5	0.28 ± 0.15	0.45 ± 0.11	1.28e-86
F_6	0.15 ± 0.08	0.33 ± 0.19	2.86e-103
F_7	0.05 ± 0.05	0.27 ± 0.19	3.08e-159

5.3.2. Peak intervals in autocorrelation of ECG

Figure 5.2 shows the box plots of features F_1 to F_7 from the peak intervals in autocorrelation of ECG. Like the previous section, feature F_3 is the only case where the two boxes are overlapping with each other. For features F_1 and F_2 , the box bottom of the class AF is close to the box top of the class Normal. For features F_4 to F_7 , boxes of the two classes are relatively far from each other, while the upper whisker of the class Normal is overlapping with the box of the class AF.



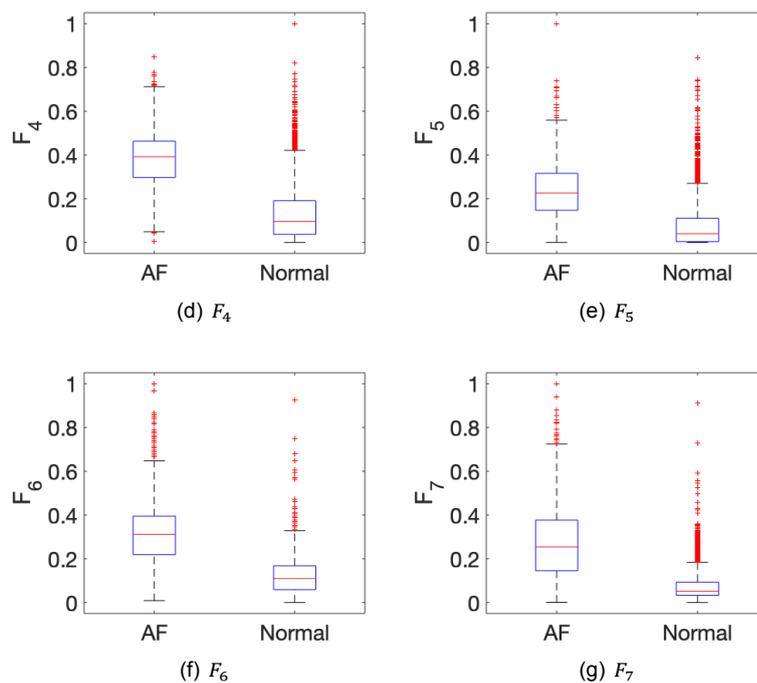


Figure 5.2: Comparison of features.

The mean value, standard deviation value, and p-value from the Mann–Whitney U test of each feature are shown in Table 5.5. The mean value of feature F_3 for the two classes is again the closest. The mean value of other features for the AF class is around twice or three times of that for the Normal class, among which the features F_4 and F_5 show the biggest difference between the two classes. All the p-values are much smaller than 0.001, meaning there is significant difference between the two classes for these features.

Table 5.5: Mean \pm standard deviation of all features

Feature	Normal	AF	p-value
F_1	0.21 ± 0.16	0.45 ± 0.13	$3.40\text{e-}121$
F_2	0.25 ± 0.18	0.49 ± 0.14	$2.43\text{e-}101$
F_3	0.20 ± 0.09	0.24 ± 0.06	$4.94\text{e-}23$
F_4	0.13 ± 0.13	0.39 ± 0.13	$1.34\text{e-}147$
F_5	0.08 ± 0.10	0.25 ± 0.14	$1.04\text{e-}123$
F_6	0.15 ± 0.09	0.34 ± 0.18	$1.05\text{e-}102$
F_7	0.09 ± 0.07	0.24 ± 0.16	$4.33\text{e-}93$

5.3.3. Peak intervals in autocorrelation of prediction error

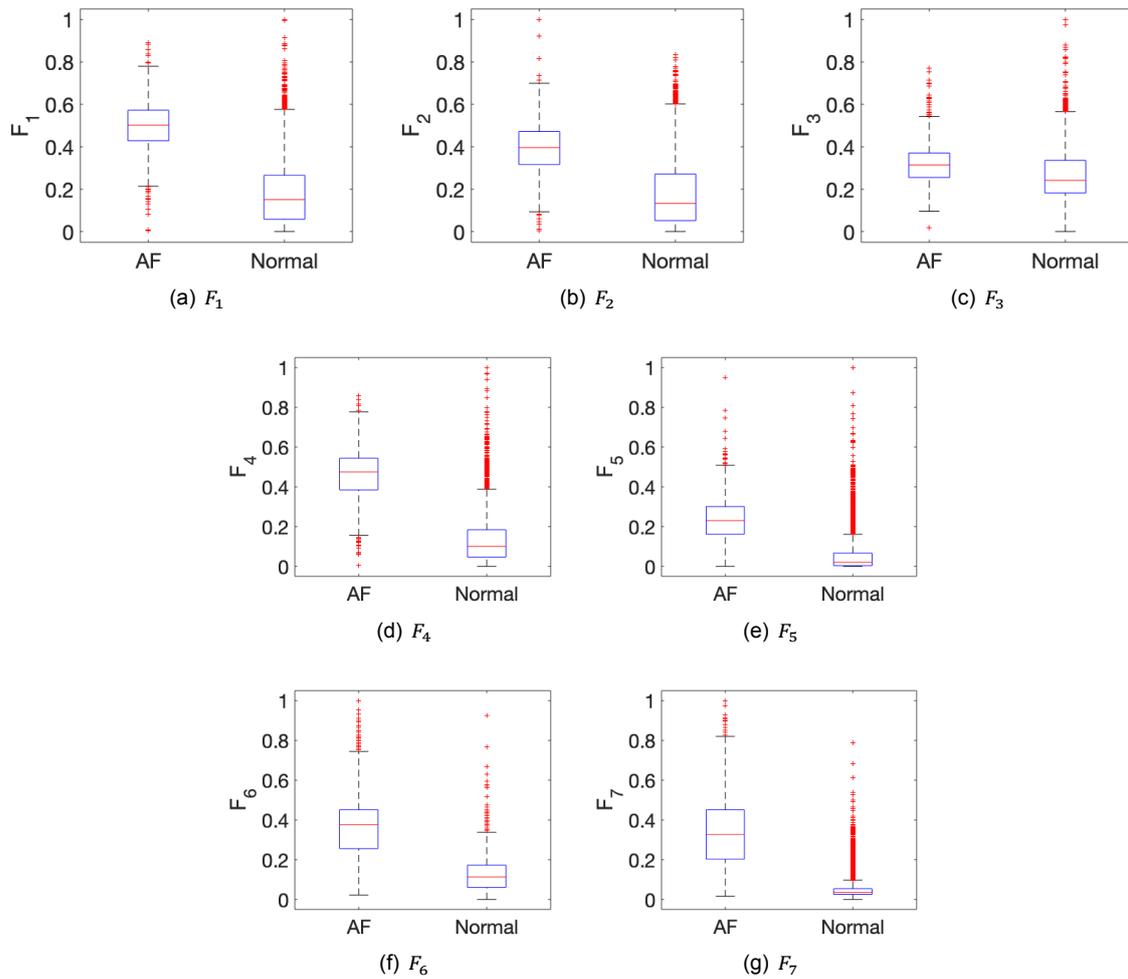


Figure 5.3: Comparison of features.

Figure 5.3 compares the value distribution of features F_1 to F_7 from the peak intervals in autocorrelation of prediction error in the form of the box plot. Feature F_3 , as usual, shows much overlapping between the two classes. Features F_1 and F_2 do not show overlapping between the boxes, but the upper whisker of the Normal class is still higher than the box of the AF class. For features F_4 and F_6 , there is no overlapping between the two boxes, but there is overlapping between the upper whisker of the Normal class and the box of the AF class. For features F_5 and F_7 , the boxes of the two classes do not overlap, and the upper whisker of the Normal class is lower than the box of the AF class.

The mean value, standard deviation value, and p-value from the Mann–Whitney U test of each feature are shown in Table 5.6. The difference of feature F_3 for the two classes is the smallest, while the difference of feature F_7 for the two classes is the biggest. All the p-values are again much smaller than 0.001, meaning there is significant difference between the two classes for these features.

Table 5.6: Mean \pm standard deviation of all features

Feature	Normal	AF	p-value
F_1	0.18 ± 0.15	0.50 ± 0.12	$2.09\text{e-}162$
F_2	0.19 ± 0.16	0.40 ± 0.13	$1.33\text{e-}106$
F_3	0.26 ± 0.11	0.32 ± 0.09	$3.93\text{e-}28$

F_4	0.14 ± 0.13	0.46 ± 0.13	2.66e-173
F_5	0.06 ± 0.09	0.24 ± 0.11	1.09e-155
F_6	0.16 ± 0.09	0.37 ± 0.20	8.38e-113
F_7	0.06 ± 0.07	0.30 ± 0.20	4.59e-146

5.4. Classification results

In this section, the training set is again selected randomly from 50% of the dataset, and the other 50% is used as the test set. To investigate the dependence of the classifier on the randomly-selected training set for different feature sets, experiments are repeated ten times, and the performance metrics are reported.

5.4.1. RR intervals

In this section, seven features including F_1 to F_7 for RR intervals are used to train both the SVM and RF classifiers. This is repeated for ten times and the mean value and the corresponding standard deviation of the performance metrics for the test set are shown in Table 5.7. For SVM, the accuracy is 93.86%, the F1 score is 0.77, and the standard deviations for these performance metrics are from 0.01 to 0.04. For RF, the accuracy is 95.46%, the F1 score is 0.81, and the standard deviations for these performance metrics are from 0 to 0.02. The F1 score for RF is slightly better than that for SVM, because the RF classifier can generalize well for the unknown data. The low standard deviation for RF is due to the large number of decision trees, which makes the majority voting insensitive for different training sets and test tests.

Table 5.7: Results

Performance	SVM	RF
Accuracy	$93.86\% \pm 0.01$	$95.46\% \pm 0$
Precision	0.75 ± 0.02	0.88 ± 0.02
Recall	0.79 ± 0.04	0.76 ± 0.02
F1 score	0.77 ± 0.02	0.81 ± 0.01

Corresponding confusion matrices for SVM and RF are given in Table 5.8 and 5.9. It can be seen from the tables that SVM correctly classifies more AF recordings than RF, at the cost of much more wrongly classified Normal rhythm. Therefore, although the F1 score for RF is higher than SVM, if the cost of false negative is much higher than false positive (in this case, it means actual AF predicted as Normal is more dangerous than actual Normal predicted as AF), it is still recommended to choose the SVM classifier.

Table 5.8: Confusion matrix

SVM	Predicted AF	Predicted Normal
Actual AF	326	53
Actual Normal	112	2425

Table 5.9: Confusion matrix

RF	Predicted AF	Predicted Normal
Actual AF	310	69
Actual Normal	45	2492

As comparison, results for only features F_1 to F_5 (features proposed in this thesis are not included) are shown below. It can be seen that when F_6 and F_7 are missing, accuracy decreases 0.02, and F1 score decreases 0.01 for SVM; while accuracy decreases 0.01, and F1 score decreases 0.03 for RF. This shows that the proposed features have extra value to the traditional Poincaré plot based features. The standard deviations for SVM increase a little, which means the performance of the classifier is not as stable as before. The standard deviations for RF are as low as the previous case, suggesting the performance of the classifier is stable even with fewer features.

Table 5.10: Results

Performance	SVM	RF
Accuracy	92.40% \pm 0.01	94.35% \pm 0
Precision	0.65 \pm 0.04	0.79 \pm 0.02
Recall	0.90 \pm 0.03	0.76 \pm 0.02
F1 score	0.76 \pm 0.04	0.78 \pm 0.01

Corresponding confusion matrices for the test set are given in Table 5.11 and 5.12. Again, SVM correctly classifies more AF class while RF correctly classifies more Normal class.

Table 5.11: Confusion matrix

SVM	Predicted AF	Predicted Normal
Actual AF	347	31
Actual Normal	149	2387

Table 5.12: Confusion matrix

RF	Predicted AF	Predicted Normal
Actual AF	303	76
Actual Normal	75	2462

5.4.2. Peak intervals in autocorrelation of ECG

In this section, seven features including F_1 to F_7 for peak intervals in autocorrelation of ECG are used to train both the SVM and RF classifiers. This is repeated for ten times and the mean value and the corresponding standard deviations of the performance metrics for the test set are shown in Table 5.13. Both Accuracy and F1 score are lower than those in the previous section. For SVM, the accuracy decreases 0.05, and the F1 score decreases 0.14. For RF, the accuracy decreases 0.02, and the F1 score decreases 0.13. This indicates that RR intervals might contain more useful information than peak intervals in autocorrelation of ECG to classify AF and Normal ECG signals. The standard deviations of SVM increases, while the standard deviations of RF stay the same. This means that with different features, the performance of SVM is more sensitive to the variation of the training set than RF.

Table 5.13: Results

Performance	SVM	RF
Accuracy	88.61% \pm 0.02	93.77% \pm 0
Precision	0.55 \pm 0.06	0.81 \pm 0.02
Recall	0.73 \pm 0.04	0.59 \pm 0.02
F1 score	0.63 \pm 0.05	0.68 \pm 0.01

Corresponding confusion matrices for the test set are given in Table 5.14 and 5.15. It is clear that in these tables, both AF and Normal are less correctly predicted than in Table 5.8 and 5.9.

Table 5.14: Confusion matrix

SVM	Predicted AF	Predicted Normal
Actual AF	283	95
Actual Normal	187	2347

Table 5.15: Confusion matrix

RF	Predicted AF	Predicted Normal
Actual AF	243	136
Actual Normal	60	2477

5.4.3. Peak intervals in autocorrelation of prediction error

In this section, seven features including F_1 to F_7 for peak intervals in autocorrelation of prediction error are used to train both the SVM and RF classifiers. This is repeated for ten times and the mean value and the corresponding standard deviations of the performance metrics for the test set are shown in Table 5.16. The results in this part are not as good as those for RR intervals, but still better than those for peak intervals in autocorrelation of ECG. This means the peak intervals in autocorrelation of prediction error could be better feature candidates than those in autocorrelation of ECG. For SVM, the standard deviations are as small as those for RR intervals; while for RF, the standard deviations increase a little compared with those for RR intervals. This suggests that the performance of SVM is as stable as before, but the performance of RF becomes sensitive to different training sets and test sets

Table 5.16: Results

Performance	SVM	RF
Accuracy	92.62% \pm 0.01	94.80% \pm 0
Precision	0.68 \pm 0.02	0.86 \pm 0.02
Recall	0.80 \pm 0.04	0.72 \pm 0.03
F1 score	0.74 \pm 0.02	0.78 \pm 0.02

Corresponding confusion matrices for the test set are given in Table 5.17 and 5.18. For SVM, comparing the table with Table 5.8, the number of correctly predicted AF data increases a little, but Normal data are less correctly predicted. For RF, both AF and Normal data are less correctly classified than those for RR intervals. Compared with Table 5.14 and 5.15, both classifiers classify more recordings for the two classes. This means when using the peak intervals in autocorrelation of prediction as features, the performance of the two classifiers are worse in general than using the RR intervals as features, but still better than using the peak intervals in autocorrelation of ECG as features.

Table 5.17: Confusion matrix

SVM	Predicted AF	Predicted Normal
Actual AF	331	48
Actual Normal	141	2395

Table 5.18: Confusion matrix

RF	Predicted AF	Predicted Normal
Actual AF	293	86
Actual Normal	57	2480

5.4.4. Combination of RR intervals and peak intervals in autocorrelation function

In this section, 21 features including F_1 to F_7 for both RR intervals and peak intervals in autocorrelation function are used to train both the SVM and RF classifiers. This is repeated for ten times and the mean value and the corresponding standard deviations of the performance metrics for the test set are shown in Table 5.19. For SVM, the accuracy is 0.95, the F1 score is 0.80, and the standard deviations are lower than previous subsections. For RF, the accuracy is 0.96, the F1 score is 0.85, and the standard deviations are also the lowest compared with other cases. Apparently, the results with all the features used are better than any of those in the previous subsections, but at the cost of higher complexity of the algorithm at the same time.

Table 5.19: Results

Performance	SVM	RF
Accuracy	94.84% \pm 0	96.37% \pm 0
Precision	0.81 \pm 0.02	0.91 \pm 0.01
Recall	0.79 \pm 0.03	0.80 \pm 0.02
F1 score	0.80 \pm 0.02	0.85 \pm 0.01

Corresponding confusion matrices for the test set are given in Table 5.20 and 5.21. It can be seen that both the classifiers show the best performance this time. While RF correctly classifies more Normal rhythms than SVM, it also show a comparable ability for classifying AF rhythm as SVM.

Table 5.20: Confusion matrix

SVM	Predicted AF	Predicted Normal
Actual AF	317	62
Actual Normal	82	2454

Table 5.21: Confusion matrix

RF	Predicted AF	Predicted Normal
Actual AF	318	61
Actual Normal	27	2510

5.5. Feature selection

Figure 5.4 shows the results of feature selection using the mrmr algorithm. The x-axis shows the number of top important features used to train the SVM classifier, and the y-axis shows the corresponding accuracy and F1 score for the test set. Each time a training set is randomly selected, causing a slight change to the ranking of the feature importance, the corresponding accuracy and F1 score for the test set are calculated. This is repeated ten times, and the final results are the average of the ten repetitions.

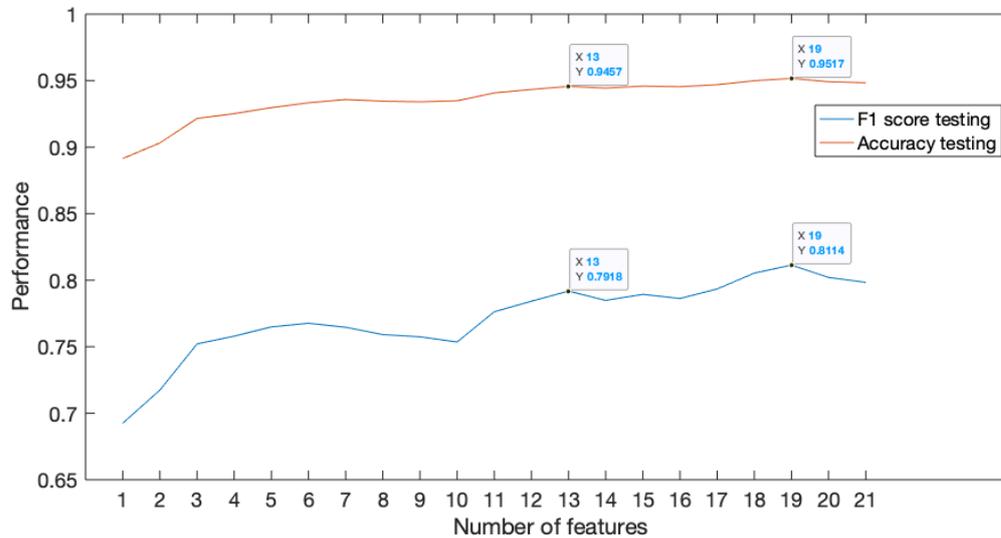


Figure 5.4: Results of feature selection.

Figure 5.4 shows that both lines are increasing more or less until the number of features is thirteen, after which they are fluctuating and do not improve too much even with more features. Therefore, although the accuracy reaches 0.95 and the F1 score reaches 0.81 with nineteen features, it is still reasonable to choose the top thirteen most important features to get a satisfying performance and reduce the complexity of the algorithm at the same time. With the top thirteen important features, the accuracy reaches 0.95, and the F1 score reaches 0.79.

For each one of the ten experiments, the most important thirteen features were recorded. The frequency of each feature ranked as the most important thirteen features for the whole ten repetitions is given in Table 5.22.

Table 5.22: Frequency of each feature showing up in the thirteen most important features during the ten experiments.

Feature	1	2	3	4	5	6	7
Frequency	9	10	10	10	10	0	10
Feature	8	9	10	11	12	13	14
Frequency	0	0	4	10	0	4	10
Feature	15	16	17	18	19	20	21
Frequency	10	0	10	10	0	3	10

From the table it can be seen that for the ten repeated experiments, eleven features including F_2 , F_3 , F_4 , F_5 , F_7 for RR intervals, F_4 , F_7 for peak intervals in autocorrelation of ECG, and F_1 , F_3 , F_4 , F_7 for peak intervals in autocorrelation of prediction error are ranked as the most important thirteen features for every single time. This implies that these eleven features contain useful information for different training sets. It can be concluded from these features that the peak intervals in autocorrelation of ECG contain less information than the other two interval series because only two features from that are included.

As for each single feature, F_6 for all the three interval series are not assumed to be important, F_1 , F_2 , F_5 for one of the interval series are selected, F_3 for two of the interval series are selected. On the other hand, F_4 and F_7 are chosen for all of the three interval series, which means they are assumed to be important and useful, and thus show their potential to classify AF and Normal ECG signals. To investigate why feature F_6 is not selected, the relationship between features F_6 and F_7 for the three interval series are illustrated in Figure 5.5.

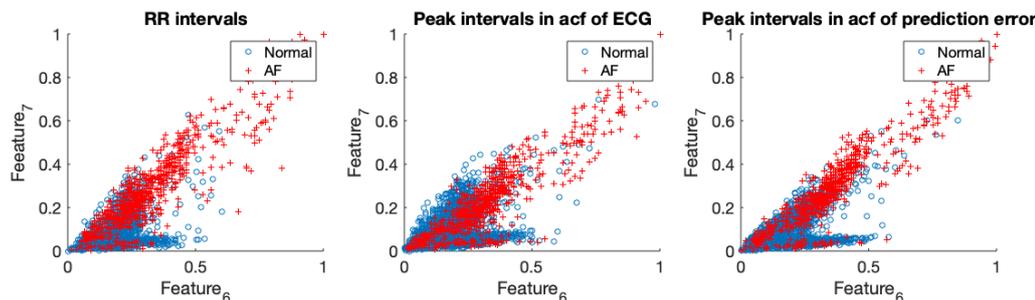


Figure 5.5: Scatter plots between features F_6 and F_7 for the three interval series. Normal rhythms are marked as blue circle. AF rhythms are marked as red plus sign.

It can be seen that these two features have a positive association. The correlation coefficient between two random variables A and B measures their random dependence and is defined as

$$\rho(A, B) = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right) \quad (5.5)$$

where μ_A and μ_B are the mean of A and B , and σ_A and σ_B are the standard deviation of A and B . The correlation coefficient between features F_6 and F_7 for the three interval series are given in Table 5.23.

Table 5.23: Correlation coefficient between features F_6 and F_7 .

Interval series	All data	Normal rhythm	AF rhythm
RR intervals	0.79	0.51	0.90
Peak intervals in acf of ECG	0.75	0.52	0.88
Peak intervals in acf of prediction error	0.81	0.55	0.92

For all the three interval series, while the correlation coefficients between the two features for Normal rhythm are around 0.5, those for AF rhythm are around 0.9, leading to those for all data being around 0.8. It can be concluded that the positive relationship between these two features are strong. Therefore, it is reasonable to choose only one feature from them.

Using the eleven most often selected features in Table 5.22 to train the SVM classifier, the accuracy and F1 score for the test test are 94.26% and 0.78, while using 21 features attains accuracy and F1 score at 94.84% and 0.80. Using the eleven most often selected features in Table 5.22 to train the RF classifier, the accuracy and F1 score for the test test are 96.15% and 0.84, while using 21 features attains accuracy and F1 score at 96.37% and 0.85. With half of the features, the F1 score of the SVM and RF classifiers decreases only 0.02 and 0.01, respectively.

5.6. M-lagged Poincaré plot analysis

Instead of plotting A_n against A_{n+1} in the Poincaré plot, some researches focus on plotting A_n against A_{n+m} to investigate the change of the plot along with the increase of lag m . It has been shown that different lags m of the Poincaré plot give a better understanding about the autonomic control of the heart rate that impact the short-term and long-term variability of the heart rate [26]. The short-term and long-term correlations can be different on different time scales. When the sampling interval is shorter than the short-term correlation length, these short-term correlations can be predominantly seen. Therefore,

the m -lagged Poincaré plot might provide more information than the conventional one-lagged Poincaré plot.

In this section, lag m increases from one to twenty, and corresponding value of feature F_7 is calculated for the whole dataset. The 25th, 75th percentile and median for the two classes are presented as the shadow area and the bold line, respectively. Figure 5.6 to 5.8 show the relationship of feature F_7 with different lag m for RR intervals, peak intervals in autocorrelation of ECG, and peak intervals in autocorrelation of prediction error, respectively.

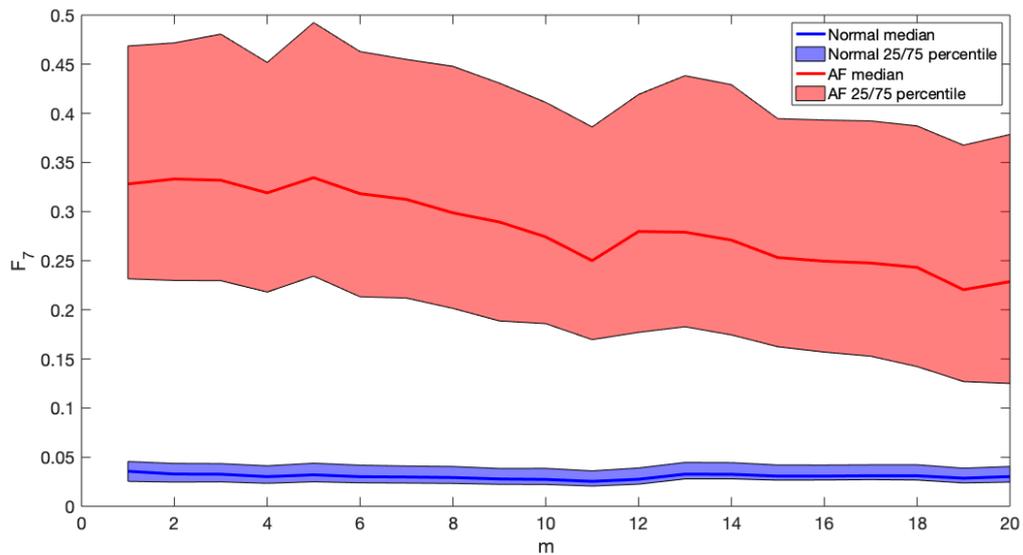


Figure 5.6: M-lagged analysis of feature F_7 for RR intervals.

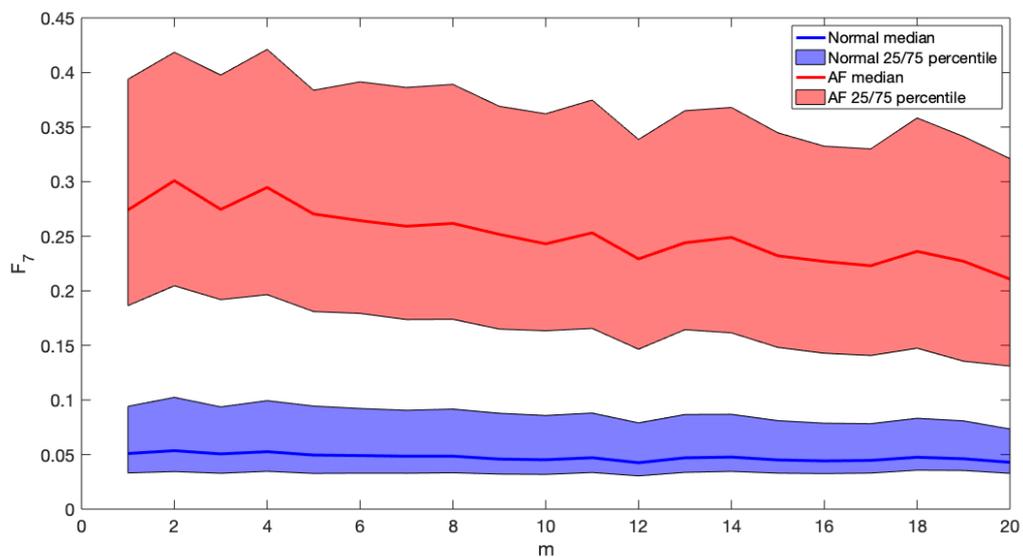


Figure 5.7: M-lagged analysis of feature F_7 for peak intervals in autocorrelation of ECG.

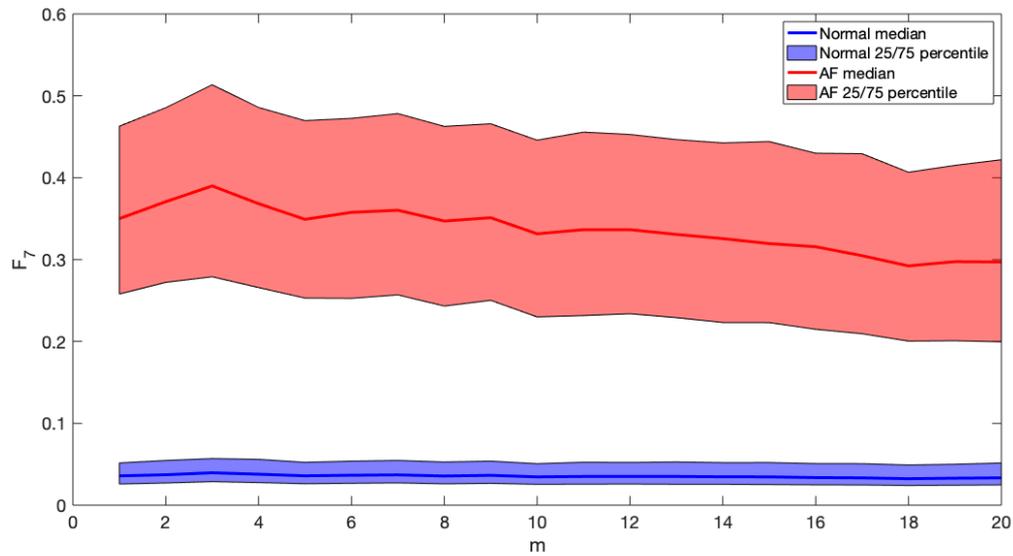
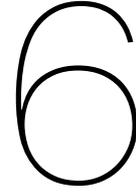


Figure 5.8: M-lagged analysis of feature F_7 for peak intervals in autocorrelation of prediction error.

It can be seen from either of the three figures that feature F_7 for the class AF has wider 25/75 percentile range than that for the class Normal, given a certain lag m . This means the variability of feature F_7 for the class AF is larger. The value for the AF rhythm is always larger than the Normal rhythm, but the gap becomes smaller as the lag m increases. So, it is better for the feature F_7 to choose a small value of the lag m , for example any number from one to four. The variation of value along the lag m for the class AF is also slightly higher than that for the class Normal. This implies that the fluctuation of feature F_7 along the increase of the lag m could also be a possible feature to classify the AF and Normal classes.



Conclusion

6.1. Conclusion

In this thesis, a new Poincaré plot based feature, which combines the distribution and position information in the Poincaré plot, is proposed to classify AF and Normal rhythms. The Poincaré plot generated from the peak intervals in the autocorrelation function in addition to the RR intervals is investigated. The performance of the new features is compared using the support vector machine and random forest classifiers.

A normal sinus rhythm ECG signal consists of P waves, QRS complexes, and T waves. For AF rhythm, the P waves are normally absent and the RR intervals become rapid and irregular. The AF period are caused by complex nonlinear processes, which can not be easily detected by the typical time domain analysis. The Poincaré plot, on the other hand, provides a visual representation of the nonlinear aspects of the heart rate variability, both qualitatively and quantitatively.

The dataset used in this thesis comes from the PhysioNet/Computing in Cardiology (CinC) Challenge 2017. After several pre-processing steps, RR intervals of the data are used to generate the Poincaré plot. Because the AF rhythm has both larger short-term and long-term variability, it shows random and chaotic patterns in the plot; while the Normal rhythm shows less clusters and usually concentrates around the center in the plot, due to its periodic characteristics.

In chapter 3, the Poincaré plot are quantized into small bins, which are given different weights by the mask processing technique. The first mask aims to approximate the distribution of the two classes, while the second mask aims to give higher weight to the states far from the center of the plot. These two masks are then element-wise multiplied and summarized with the quantized Poincaré plot of the new data, resulting in a higher value if the new data is AF, or the opposite if the new data is Normal. By doing so, the new feature aims to approximate the expected value of the matrix of the Poincaré plot, whose entries are ones if corresponding states are visited by the system, otherwise are zeros. Meanwhile, It attempts to increase the gap between the two classes by giving more weight to the outliers, which are more likely to occur during an AF period.

Another contribution in chapter 3 is that the peak intervals in the autocorrelation function of both ECG and prediction error are used to generate the Poincaré plot, similar as the typical Poincaré plot does with the RR intervals. The autocorrelation function evaluates the self-periodicity of a time series, and thus is assumed to add extra information to the heart rate variability. Therefore, some commonly used Poincaré plot based features together with the new feature proposed in this thesis, are extracted from the Poincaré plot based on both the RR intervals and the peak intervals of the autocorrelation function. 21 features in total are included in the thesis, including three standard deviation related features, two temporal based features, and two features that are investigated in this thesis.

In chapter 4, 21 features are ranked based on their importance calculated by the minimum redundancy maximum relevance algorithm. In general, the RR interval related features contain the most useful information, followed by the peak intervals in the autocorrelation of prediction error related features. Peak intervals in the autocorrelation of ECG related features are ranked as the least important. Classifiers used during the experiments include SVM and RF. SVM has been used widely to classify nonlinear data by mapping them into higher dimensions using the kernel trick. The computational com-

plexity of the algorithm depends on the number of support vectors instead of the number of the training data, therefore reduces the curse of dimensionality to some extent and increases the robustness of the model. Random forests are an ensemble learning method consisting of a collection of decision trees. The randomness between various decision trees originates from the randomly selected input data and feature set for each tree. This makes the random forest generalize well to the unknown data.

Chapter 5 presents the results of the experiments. The comparison of the distribution of the feature values in the box plot shows that the proposed feature has no overlapping between the two boxes for the two classes, for any of the three kinds of intervals. As for the performance of the two classifiers, using all these 21 features generates the highest F1 score of 0.80 and 0.85 for SVM and RF, respectively. When using only seven features from the same intervals, RR intervals generates the highest F1 score of 0.77 and 0.81 for SVM and RF, followed by the peak intervals in the autocorrelation of prediction error at 0.74 and 0.78 for SVM and RF, and the peak intervals in the autocorrelation of ECG gives the lowest F1 score of 0.63 and 0.68 for SVM and RF. The poor performance of the peak intervals in the autocorrelation of ECG is probably due to the definition of the peak in the autocorrelation function. Only looking for peaks that are higher than a threshold and far from each other with a certain distance might not be enough to exploit the autocorrelation function's potential of evaluating the self-similarity.

For the above results, random forests always achieve a higher F1 score with a lower variance than SVM for ten repeated experiments. This is mainly because of the large number of trees in the forest and the majority vote among various trees that makes the forest avoid overfitting and perform stably to different training set and test set. Using the mrmr feature selection algorithm, eleven most important features are selected and lead to a comparable F1 score of 0.78 and 0.84 for SVM and RF, respectively.

6.2. Future work

This thesis achieves a decent performance to classify AF and Normal ECG signals using only the Poincaré plot based features. However, when checking the confusion matrix of both the SVM or RF, more proportion of AF signals are wrongly predicted as Normal signals than vice versa. This means that a number of AF rhythms do not show noticeable dispersion in the Poincaré plot. This leads to the other characteristics of the AF rhythms: the absence of the P waves. Other features that investigate the morphology of the ECG signals are recommended to be considered to further improve the accuracy of the classifiers.

In addition to normal and AF rhythms, the dataset used in this thesis also includes other arrhythmia and noisy classes, which are found to be difficult to differentiate by many teams during the CinC challenge 2017. The other arrhythmia contains relatively regular RR intervals similar to the normal class, and the noisy class shows a chaotic pattern as the AF class does. This makes it difficult to classify all the four classes, and using only the Poincaré plot based features is not enough in this case. Therefore, time-frequency features like discrete wavelet transform (DWT) features and other nonlinear analysis are suggested to be investigated.

Deep learning algorithms appear to be promising these days and therefore might be helpful with this problem. They do not need the feature extraction and reach a satisfying performance as long as the parameters and the structure of the model are well tuned. One limitation of deep learning algorithms is that they are lacking interpretation.

One possible application of detecting arrhythmia is to implement the algorithm in the wearable devices, in which case the requirements include real-time detection, low complexity, and low power consumption. Therefore, the balance between the accuracy rate and the complexity of the algorithm and the size of the feature space needs to be found. One example are deep learning methods, which could generate a high performance but are computationally expensive at the same time.

Bibliography

- [1] Why Atrial Fibrillation (AF or AFib) Matters. <https://www.heart.org/en/health-topics/atrial-fibrillation/why-atrial-fibrillation-af-or-afib-matters>, . Online; accessed 4-July-2020.
- [2] What are the Symptoms of Atrial Fibrillation (AFib or AF)? <https://www.heart.org/en/health-topics/atrial-fibrillation/what-are-the-symptoms-of-atrial-fibrillation-afib-or-af>, . Online; accessed 4-July-2020.
- [3] Atrial Fibrillation in Children. <https://www.heart.org/en/health-topics/atrial-fibrillation/who-is-at-risk-for-atrial-fibrillation-af-or-afib/atrial-fibrillation-in-children>, . Online; accessed 4-July-2020.
- [4] visualize summary statistics with box plot - matlab boxplot - mathworks benelux. <https://nl.mathworks.com/help/stats/boxplot.html>. Online; accessed 25-June-2020.
- [5] Electrocardiogram (EKG, ECG). <https://www.cvphysiology.com/Arrhythmias/A009>. Online; accessed 21-Aug-2020.
- [6] A Beginner's Guide To Understanding Convolutional Neural Networks. <https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/>. Online; accessed 23-May-2020.
- [7] Heart rate: What is a normal heart rate? <https://www.medicalnewstoday.com/articles/235710>. Online; accessed 16-May-2020.
- [8] Decision Trees and Random Forests. <https://towardsdatascience.com/decision-trees-and-random-forests-df0c3123f991>. Online; accessed 19-Aug-2020.
- [9] Lilliefors test - MATLAB lillietest - MathWorks Benelux. <https://nl.mathworks.com/help/stats/lillietest.html>. Online; accessed 16-July-2020.
- [10] Rank features for classification using minimum redundancy maximum relevance (MRMR) algorithm - MATLAB fscmrmr - MathWorks Benelux. <https://nl.mathworks.com/help/stats/fscmrmr.html>. Online; accessed 2-June-2020.
- [11] Cross-correlation - MATLAB xcorr - MathWorks Benelux. <https://nl.mathworks.com/help/matlab/ref/xcorr.html>. Online; accessed 26-May-2020.
- [12] Atrial fibrillation. <https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624>, Jun 2019. Online; accessed 4-July-2020.
- [13] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [14] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE transactions on biomedical engineering*, 48(11):1342–1347, 2001.
- [15] Gari D Clifford, Chengyu Liu, Benjamin Moody, H Lehman Li-wei, Ikaro Silva, Qiao Li, AE Johnson, and Roger G Mark. Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE, 2017.
- [16] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

- [17] Shreyasi Datta, Chetanya Puri, Ayan Mukherjee, Rohan Banerjee, Anirban Dutta Choudhury, Rituraj Singh, Arijit Ukil, Soma Bandyopadhyay, Arpan Pal, and Sundeep Khandelwal. Identifying normal, af and other abnormal ecg rhythms using a cascaded binary classifier. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE, 2017.
- [18] Zümray Dokur and Tamer Ölmez. Ecg beat classification by a novel hybrid neural network. *Computer methods and programs in biomedicine*, 66(2-3):167–181, 2001.
- [19] JP Eckmann, S Oliffson Kamphorst, D Ruelle, et al. Recurrence plots of dynamical systems. *World Scientific Series on Nonlinear Science Series A*, 16:441–446, 1995.
- [20] Masatake Fukunami, Takahisa Yamada, Masaharu Ohmori, Kazuaki Kumagai, Kiyoshi Umemoto, Akihiko Sakai, Nobuhiko Kondoh, Tetsuo Minamino, and Noritake Hoki. Detection of patients at risk for paroxysmal atrial fibrillation during sinus rhythm by p wave-triggered signal-averaged electrocardiogram. *Circulation*, 83(1):162–169, 1991.
- [21] İnan Güler and Elif Derya Übeyli. Ecg beat classifier designed by combined neural network model. *Pattern recognition*, 38(2):199–208, 2005.
- [22] Yuki Hagiwara, Hamido Fujita, Shu Lih Oh, Jen Hong Tan, Ru San Tan, Edward J Ciaccio, and U Rajendra Acharya. Computer-aided diagnosis of atrial fibrillation based on ecg signals: A review. *Information Sciences*, 467:99–114, 2018.
- [23] Shenda Hong, Meng Wu, Yuxi Zhou, Qingyun Wang, Junyuan Shang, Hongyan Li, and Junqing Xie. Encase: An ensemble classifier for ecg classification using expert features and deep neural networks. In *2017 Computing in cardiology (cinc)*, pages 1–4. IEEE, 2017.
- [24] Che-Hao Hsu, Ming-Ya Tsai, Go-Shine Huang, Tso-Chou Lin, Kuen-Pao Chen, Shung-Tai Ho, Liang-Yu Shyu, and Chi-Yuan Li. Poincaré plot indexes of heart rate variability detect dynamic autonomic modulation during general anesthesia induction. *Acta Anaesthesiologica Taiwanica*, 50(1):12–18, 2012.
- [25] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification, 2003.
- [26] Chandan K Karmakar, Ahsan H Khandoker, Jayavardhana Gubbi, and Marimuthu Palaniswami. Complex correlation measure: a novel descriptor for poincaré plot. *Biomedical engineering online*, 8(1):17, 2009.
- [27] Gustavo Lenis, Nicolas Pilia, Axel Loewe, Walther HW Schulze, and Olaf Dössel. Comparison of baseline wander removal techniques considering the preservation of st changes in the ischemic ecg: a simulation study. *Computational and mathematical methods in medicine*, 2017, 2017.
- [28] Chun-Cheng Lin and Chun-Min Yang. Heartbeat classification using normalized rr intervals and morphological features. *Mathematical Problems in Engineering*, 2014, 2014.
- [29] Ali R Mani, Sara Montagnese, Clive D Jackson, Christopher W Jenkins, Ian M Head, Robert C Stephens, Kevin P Moore, and Marsha Y Morgan. Decreased heart rate variability in patients with cirrhosis relates to the presence and degree of hepatic encephalopathy. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 296(2):G330–G338, 2009.
- [30] Jonathan Moeyersons, Matthew Amoni, Sabine Van Huffel, Rik Willems, and Carolina Varon. R-deco: An open-source matlab based graphical user interface for the detection and correction of r-peaks. *PeerJ Computer Science*, 5:e226, 2019.
- [31] Chrysostomos L Nikias and Jerry M Mendel. Signal processing with higher-order spectra. *IEEE Signal processing magazine*, 10(3):10–37, 1993.
- [32] Saurabh Pal and Madhuchhanda Mitra. Empirical mode decomposition based ecg enhancement and qrs detection. *Computers in biology and medicine*, 42(1):83–92, 2012.

- [33] Jinho Park, Sangwook Lee, and Moongu Jeon. Atrial fibrillation detection by heart rate variability in poincare plot. *Biomedical engineering online*, 8(1):38, 2009.
- [34] B Pourbabaee and C Lucas. Automatic detection and prediction of paroxysmal atrial fibrillation based on analyzing ecg signal feature classification methods. In *2008 Cairo International Biomedical Engineering Conference*, pages 1–4. IEEE, 2008.
- [35] Pratiksha Sarma, SR Nirmala, and Kandarpa Kumar Sarma. Classification of ecg using some novel features. In *2013 1st International Conference on Emerging Trends and Applications in Computer Science*, pages 187–191. IEEE, 2013.
- [36] Reem Satti, Noor-UI-Hoda Abid, Matteo Bottaro, Michele De Rui, Maria Garrido, Mohammad Reza Rauofy, Sara Montagnese, and Alireza Mani. The application of the extended poincaré plot in the analysis of physiological variabilities. *Frontiers in physiology*, 10:116, 2019.
- [37] Koo Ping Shung. Accuracy, Precision, Recall or F1? <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>, Apr 2020. Online; accessed 24-June-2020.
- [38] Tomás Teijeiro, Constantino A García, Daniel Castro, and Paulo Félix. Arrhythmia classification from the abductive interpretation of short single-lead ecg records. In *2017 Computing in cardiology (cinc)*, pages 1–4. IEEE, 2017.
- [39] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern recognition & Matlab intro*. Academic Press, Inc., 2010.
- [40] Mikko P Tulppo, TH Makikallio, TE Takala, THHV Seppanen, and Heikki V Huikuri. Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American journal of physiology-heart and circulatory physiology*, 271(1):H244–H252, 1996.
- [41] Developed with the special contribution of the European Heart Rhythm Association (EHRA), Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS), Authors/Task Force Members, A John Camm, Paulus Kirchhof, Gregory YH Lip, Ulrich Schotten, Irene Savelieva, Sabine Ernst, Isabelle C Van Gelder, et al. Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the european society of cardiology (esc). *European heart journal*, 31(19):2369–2429, 2010.
- [42] Can Ye, Miguel Tavares Coimbra, and BVK Vijaya Kumar. Arrhythmia detection and classification using morphological and dynamic features of ecg signals. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 1918–1921. IEEE, 2010.
- [43] Morteza Zabihi, Ali Bahrami Rad, Aggelos K Katsaggelos, Serkan Kiranyaz, Susanna Narkilahti, and Moncef Gabbouj. Detection of atrial fibrillation in ecg hand-held devices using a random forest classifier. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE, 2017.
- [44] Zhenyu Zhao, Radhika Anand, and Mallory Wang. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. *arXiv preprint arXiv:1908.05376*, 2019.