Automatic Atrial Fibrillation Detection Based on Heart Rate Variability and Spectral Features

Seminar Report

*Submitted in partial fulfillment of the requirements for
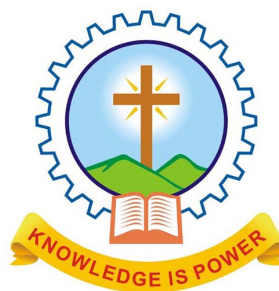the award of degree of*

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

MUHAMMED RASHID K K

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

Automatic Atrial Fibrillation Detection Based on Heart Rate Variability and Spectral Features

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

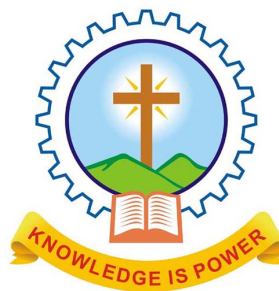**BACHELOR OF TECHNOLOGY**

In

COMPUTER SCIENCE AND ENGINEERING

of

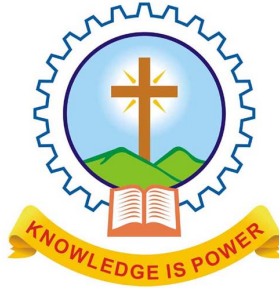**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**

Submitted By

MUHAMMED RASHID K K

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the report entitled **Voice Disorder Identification by using Machine Learning Techniques** *submitted by Mr. MUHAMMED RASHID K K S, Reg. No. MAC15CS041 towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for the year 2019 is a bonafied record of the seminar carried out by him under our supervision and guidance.*

....
Prof. Joby George
Faculty Guide

....
Prof. Neethu Subash
Faculty Guide

....
Dr. Surekha Mariam Varghese
Head of the Department

Date:

Dept. Seal

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the God Almighty for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Solly George, Principal and Dr. Surekha Mariam Varghese, Head Of the Department for providing the necessary facilities and their encouragement and support.

I owe special thanks to the staff-in-charge Prof. Joby george, Prof. Neethu Subash and Prof. Joby Anu Mathew for their corrections, suggestions and sincere efforts to co-ordinate the seminar under a tight schedule.

I express my sincere thanks to staff members in the Department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally, I would like to acknowledge the heartfelt efforts, comments, criticisms, co-operation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish.

ABSTRACT

Atrial brillation is one of the most common sustained arrhythmias .It affects about 1% of the population around the world. Rapid popularization of portable and wearable devices has made widespread personalized and mobile health-care get closer to reality than ever before. It proposes a method aiming for automatic detection of AF from short single lead electrocardiogram recordings. AF is a kind of arrhythmia that alters the dynamics of heart rhythms.The morphological characteristics in ECG tracings,heart rate variability analysis are adopted as feature extractors. The method is validated on a public available dataset comprised of short ECG recordings of normal rhythm, AF, and other arrhythmias by support vector machine and baggingtrees. For two-class classication problems accuracy varies from 92.0% to 96.0% under different additional noise levels.For three-class classication problem accuracy as high as 82.0% is obtained. Experimental results suggest that even for a relatively short ECG recording, non-linear descriptors of HRV are still efcient and robust for AF detection.

Contents

List of Figures

LIST OF ABBREVIATION

AF   Atrial Fibrillation

ECG   Electrocardiogram

HRV   Heart Rate Variability

CW   Characteristic Wave

AAFD   Automatic Atrial Fibrillation Detection

RRI   R-peak Intervals

PSD   Power Spectrum Density

ANS   Autonomous Nervous System

CCM   Complex Correlation Measure

MRMR   Maximum Relevance and Minimum Redundancy

RV   Relevance

RD   Redundancy

CHAPTER 1

Introduction

Atrial fibrillation (AF) is one of the most common sustained arrhythmias in the world with the elderly being more likely to be affected. It has been reported that more than 5% of people older than 65 suffer from this disorder. Men are more vulnerable to AF than women [1]. AF may lead to stoke and heart failure, so timely diagnosis and control are crucial in the management of AF.

Atrial fibrillation's can be classified into paroxysmal, persistent and permanent according to the medical history of patients and whether it can be terminated spontaneously or by external interventions. Although an abnormal tissue as substrate is identified as an important factor maintaining the arrhythmia, AF is also observed among patients without any structural heart disease. After persistent atrial fibrillation, the atrium of patients could adapt to this pathological condition in the aspects of ion channel functions of cardiomyocyte, called remodeling. The shortened refractory period (RP) of the action potential (AP) is a typical electrophysiological alteration caused by AF leads to tachycardia and possibly the absence of P waves, which can be rejected through electrocardiogram (ECG) tracing. And thus ECG is a widely used tool for the diagnosis of AF.

The ultimate goal of the treatment of AF is the restoration of sinus rhythm. Pharmacological cardioversions and external/ internal electrical cardioversions are efficient for recent onset of AF and persistent AF respectively. For AF initiated by some foci, for example, pulmonary vein, known as being responsible for the onset of left atrial tachycardia, radio frequency ablation causes the necrosis of specific region so the tissue in this region will not be electrically active again. In addition, surgery and implantation of pacemaker are also options. As for the diagnosis of AF, detection of abnormal episodes in ECG recordings provides useful information to clinicians and is usually served as the first step of the diagnosis procedure. In clinical

practice, experienced cardiologists inspect ECG tracing manually, which is a time-consuming work.For long-term recording acquired by holter monitors, it is especially hard for cardiologists to locate the abnormal episodes. So it is an urgent need to develop efficient and robust decision support systems for automatic detection of atrial fibrillation (ADAF).

Most of these kinds of decision support systems take ECG signals as inputs. Then an algorithm is performed to translate the raw data into useful features and finally a graspable result for clinicians is outputted. Different kinds of features inspired by medical knowledge or driven by data analytics have been adopted in related research. Among them, the abnormal P wave and QRS morphology are widely validated by clinicians as indicative signs of AF with physiological interpret-ability. Instead of encoding the medical knowledge explicitly, features of a data-driven nature are also adopted. For example, inter-pulse intervals (R-R intervals, RRI), defined as the durations between consecutive R peaks, code information about not only coarse-grained heart rate (HR) as well as dynamics characteristics of heart activities. RRI based features, including heart rate variability (HRV), are investigated in detail for ADAF. Furthermore, frequency spectrum analysis and nonlinear dynamics analysis about both raw data and some discrete time series like RR intervals are also proposed to be effective feature extraction methods. These features are usually not intuitionally like the morphology of characteristic waves (CW) in ECG.

Quantitative metrics are indispensable before such kind of information entering into medical context. All the automatic detection methods mentioned above are based on the recording and analysis of ECG signal. Holter monitoring and implantable pacemaker capable of recording and storing the ECG signals have been adopted for many years. However, the application of such technologies is still limited because the normal use of them is, to some extent, obtrusive. The prevalence of wearable and portable devices aiming for continuous and unobtrusive bio-signal monitoring and health supports has been witnessed during recent years. For example, many wearable and portable ECG monitoring systems have been proposed. The interfaces between the system and human body are diversified. Most of the interfaces are of contact-type. Except for pre-gelled Ag/Ag-Cl electrode, other kind of electrodes made of textile, polymer materials, are reported. Non-contact ECG sensing, mainly based on capacitive

coupling, are also studied. The skin-to-electrode interface can make significant impact on the quality of acquired signal. Camera based and bio-radar based modalities have also been investigated to provide alternatives to reject related information about heart dynamics.

CHAPTER 2

Existing methods

## 2.1 Atrial Fibrillation Detection Using an iPhone 4S

An iPhone App was developed using the Objective-C pro- gramming language. For the pulsatile signal acquisition, the iPhone 4S videos were recorded, and the signal was obtained by averaging 50 50 pixels of the green band for every frame [8], [11]. The sampling rate for iPhone 4s was 30 frames/s. How- ever, in rare cases, the sampling rate was slightly lower (e.g., 25 Hz due to internal processing load). Due to the frame rate variability, we interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. The peak detection algorithm incorporated a filter bank with variable cut- off frequencies, spectral estimates of the heart rate, rank-order nonlinear filters and decision logic [12]. The final result shows either NORMAL or AF DETECTED.

2.1.1 Statistical Approach for AF Detection

The RMSSD is used to quantify beat-to-beat variability. Since AF exhibits higher variability than NSR, the RMSSD is expected to be higher than those of NSR RR time series. As subjects have different mean heart rates, we normalize by dividing the RMSSD by the mean value of the RR time series.

The second component of the AF detection algorithm isShE. The ShE provides a quantitative measure of uncertainty for a random variable. For example, a random white noise signal is expected to have the highest ShE value due to maximum uncertainty in predicting the pat-

terns of the signal. Fig. 2.1 shows a histogram of the ShE for AF (combined MIT-BIH AF and iPhone AF) and NSR segments from the MIT-BIH NSR and iPhone NSR databases. Similar to RMSSD, ShE values for AF are higher than NSR, and a good separation between them is observed.
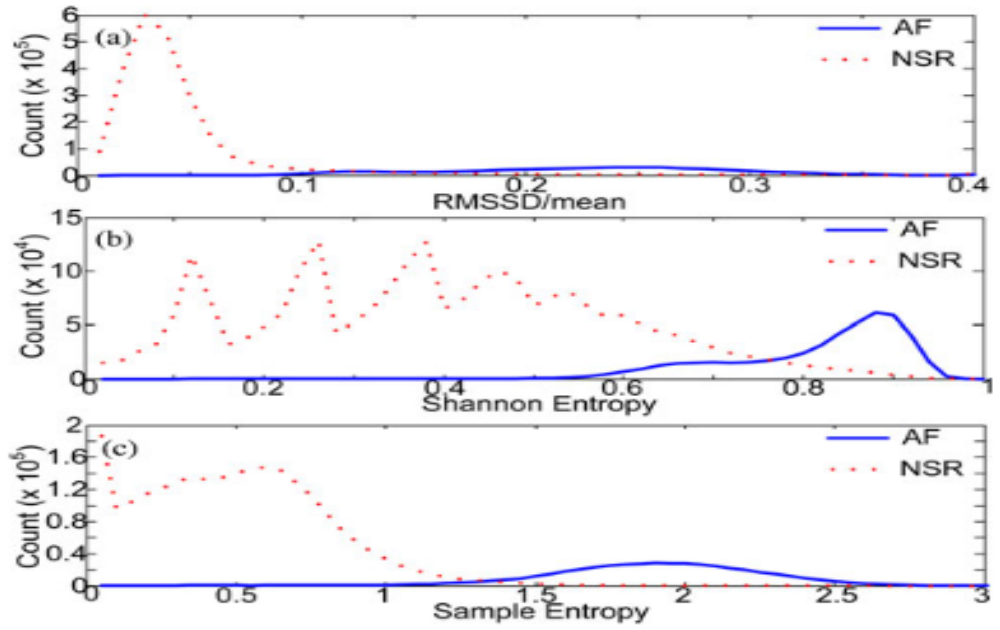
.png

Fig. 2.1: (a) RMSSD/mean values, (b) ShE values, and (c) SampE values.

The third component of the AF detection algorithm is the- SampE. The SampE is the negative natural logarithm of an esti- mate of the conditional probability that a subseries that matches pointwise within a tolerance r also match at the next point, where self-matches are not included in calculating the probability. A high value of SampE indicates low similarity in the time series while a low value of Sample entropy indicates high similarity. Thus, the SampE is a useful tool to assess randomness of RR time series. Fig. 2.1(c) shows a histogram of the SampE for AF (combined MIT-BIH AF and iPhone AF) and NSR segments from MIT-BIH AF and iPhone NSR databases. The SampE val- ues are higher for AF than NSR, and similar to RMSSD and ShE calculations, a good separation between AF and NSR is observed.

CHAPTER 3

Proposed method

3.1 Materials and methods

Our aim is to develop a robust AAFD method processing data collected from wearable or portable devices in non-structured environments. Data collected from such devices are usually vulnerable to different kinds of artifacts. A relatively long and clean section of signal cannot always be anticipated from wearable systems because serious artifacts could scatter among continuous recording and thus interrupt available signal into smaller slices. Portable devices, on the other hand, do not always work in a wearable fashion or attached to the patients. The length of recording depends on the users' behavior so is unmanageable for the developers. Therefore, morphological characteristics including QRS complex length, PR/ST intervals etc. even being clinically relevant and physiologically sound, are not preferred because these subtle features cannot always be guaranteed in the scenarios mentioned above.

However, R peak, the most dominant fiducial point in ECG signal, is usually the most unaffected feature in ECG. If the R peaks are not identifiable in a segment of ECG, the other fiducial points or CWs are more unlikely to provide useful information then. On the contrary, when other fiducial points or CWs are not available, R peak is still possible to be detected with high confidence and accuracy, typically happening under low or moderate noise level. Therefore, heart rate variability (HRV) based features are considered for their robustness under noisy scenarios. In addition, HRV based features code only dynamics characteristics about heart activities. Detailed information carried by ECG signals, although vulnerable in time domain, can be preserved partially in frequency domain. In this, spectral features are adopted

6

as supplementary.

3.2   Electrocardiogram

Electrocardiography (ECG or EKG) is the process of recording the electrical activity of the heart[4] over a period of time using electrodes placed over the skin. These electrodes detect the tiny electrical changes on the skin that arise from the heart muscle's electrophysiologic pattern of depolarizing and repolarizing during each heartbeat. It is very commonly performed to detect any cardiac problems.

In a conventional 12-lead ECG, ten electrodes are placed on the patient's limbs and on the surface of the chest. The overall magnitude of the heart's electrical potential is then measured from twelve different angles ("leads") and is recorded over a period of time (usually ten seconds). In this way, the overall magnitude and direction of the heart's electrical depolarization is captured at each moment throughout the cardiac cycle.[5] The graph of voltage versus time produced by this noninvasive medical procedure is an electrocardiogram.

During each heartbeat, a healthy heart has an orderly progression of depolarization that starts with pacemaker cells in the sinoatrial node, spreads throughout the atrium, passes through the atrioventricular node down into the bundle of His and into the Purkinje fibers, spreading down and to the left throughout the ventricles.[6] This orderly pattern of depolarization gives rise to the characteristic ECG tracing. To the trained clinician, an ECG conveys a large amount of information about the structure of the heart and the function of its electrical conduction system.[7] Among other things, an ECG can be used to measure the rate and rhythm of heartbeats, the size and position of the heart chambers, the presence of any damage to the heart's muscle cells or conduction system, the effects of heart drugs, and the function of implanted pacemakers.[8]

3.3   Data description

In this work, a public available dataset released as the material for Computing in Cardiology Challenge (CinC) 2017 is used. The public part consists of 8528 single lead ECG recordings with length varying from 9 seconds to over 1 minute. The labeling of the recordings is performed by 4 clinical cardiologists independently. About 60% of the data are classified as normal rhythm (N), 9% of the data are diagnosed as episodes of atrial fibrillation (A). About 28% of the data belong to arrhythmias other than AF (O), with no more detailed information provided. Finally, about 3% of the data are identified as too noisy to be useful.

Class	Numbers of recordings	Portion
Normal Rhythm(N)	5076	59.52%
Other Arrhythmias(O)	2415	28.32%
Atrial Fibrillation (A)	758	8.89%
Noisy (~)	279	3.27%

.png

Fig. 3.1: Class distribution of the dataset

3.4   Feature extraction

3.4.1   HRV based features

R peak detection is the first step to extract HRV based features. A famous real-time QRS detection algorithm developed by Pan and Tompkins (PT) is adopted for its superior performance. The indices of R peaks in ECG recording are identified. Denote the number of R peaks detected as N, the sampling frequency of the signal as f Hz and the length of this
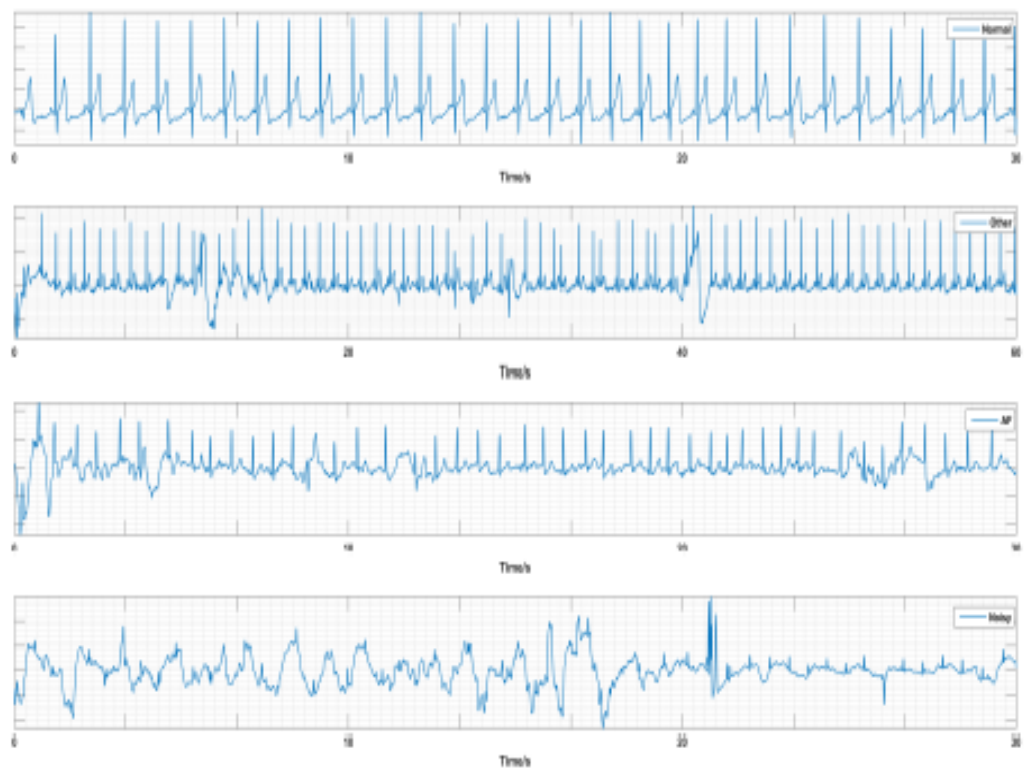
.png

Fig. 3.2: Sample recordings from each class

recording as L. The average HR (beats/min) is obtained as HR $= 60$ x (N f / L) And a sequence RRI comprised of RRI in order: RRI $=$: [RR1,RR2,....RR(N-1)]

.png
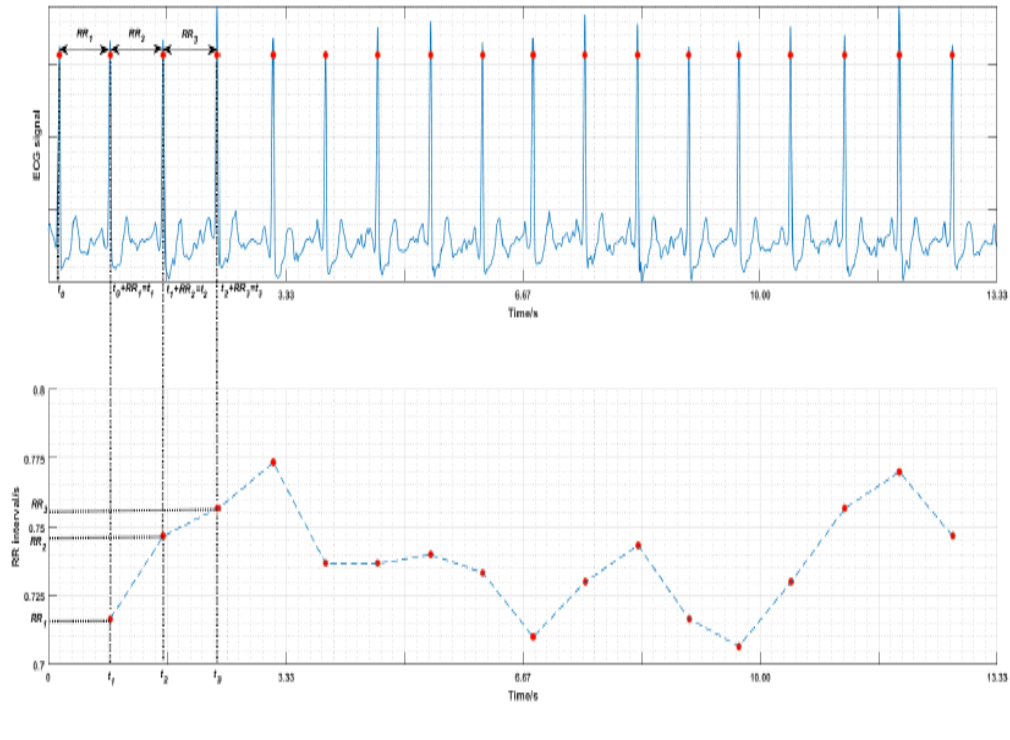
Fig. 3.3: HRV from ECG: (a) Original signal. (b) Inter pulse intervals as an unevenly spaced signal

Heart rate variability (HRV) is the oscillations between consecutive heart beat intervals. HRV was not regarded as a digital signal, for all intents and purposes, when it was observed and conceptualized in medical context for the first time. However, frequency analysis was introduced later for the relation between the activities of autonomous nervous system (ANS) and the energy distribution of HRV in frequency domain. HRV is believed to be modulated by both sympathetic and vagal nervous systems. So the power spectrum density (PSD) of HRV can be interpreted in terms of sympathetic-vagal balance. However there are some intrinsic flaws for HRV being evaluated as a digital signal sampled in some manner.

a

ECG → BPF → $x[n]$ → Differentiation → $v_e[n]$ → Transform → $y[n]$

Compare to Threshold → If appropriate then apply secondary threshold → Search for real peak → Indices of R-peaks

b

ECG → BPF → Differentiation → $v_s[n]^\gamma$ → Time-averaging

Compare to threshold → If RRI[k]>1.5*RRI[k-1] then apply secondary threshold → Search for real peak → Indices of R-peaks
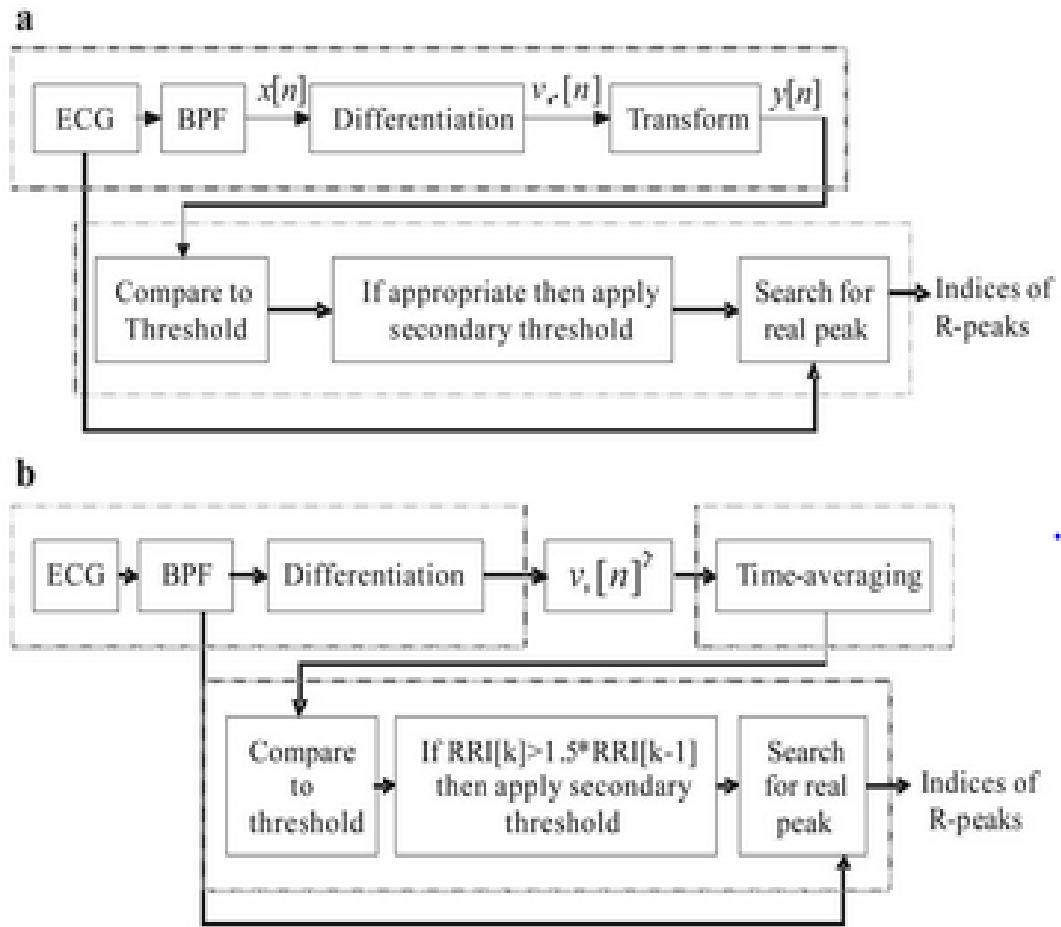
.png

Fig. 3.4: QRS detection algorithm

A typical RRI sequence as a function about time and the original ECG from which the RRI sequence is extracted are provided in Figure 3. Only after the next R peak is identified, the last RRI can thus be determined, and if the human hearts beat in a strict steady rhythm, there will be no HRV. The RRI sequence, if being taken as a digital signal, is irregularly spaced in time with dimension of quantity being also the time instead of mass of length etc. Interpolation and re-sampling or Lomb-Scargle Periodogram are frequently used techniques to circumvent this unevenness in time domain. However, these methods assume the existence of a hypothetical continuous random process from which any evenly or unevenly spaced signal can be sampled depending on our choice. Therefore, subtle side effects are introduced by different interpolation criteria. Of course, the evenly spaced fashion is welcomed under almost all circumstances and no additional operations are necessary but it should be noticed that the existence of such a continuous random process is, although intuitive in many fields including but not limited to astro-observations, not self-evident for HRV. When we want to take it as a function about time, it becomes a peculiar quantity quantified in the unit of time while also varying as time elapses. A better model could be point process, which circumvent the explicit or implicit interpolation operations among existing methods in form by providing parametric probability distributions whereby quantities of interest can be calculated analytically from the formula with estimated coef- fcients. However, the advantages of point process based heart rate variability analysis is still fatalistic because no brand new metrics has been derived from it and the unsatisfactory interpolation was just replaced with a more subtle assumption instead of truly removed.

Except for the conceptual difficulties, an acceptable frequency resolution could not always be satisfied in wearable scenarios. Typical value of length of recording is significantly lower than that was recommended by EHA.

Meanwhile, the counterpart of the famous uncertainty principle of quantum physics in signal processing asserts the time resolution and frequency resolution of a given signal cannot be arbitrarily high concurrently. Given a signal s(t) with normalized energy distribution on time axis

$\int_{-\infty}^{\infty}(S(t))^2 dt = 1$

$(\int_R (S(t))^2 dt)(\int_R f^2 (S(f))^2 df) \leq 1$

where s(f) is the Fourier transform of s(t). For signals with very compact supports in time domain like RRI sequences obtained here, the energy concentrations of the corresponding signal in frequency domain is poor.

According to the reasons mentioned above, direct power spectrum density estimation is waived in this research. Instead, RRI sequence and its first order difference are arranged into patternized matrix and mapped into 2-D plane (Poincare Plot) and then several descriptors about the patternized matrix and Poincare Plot are supposed to be able to characterize the detailed dynamical information carried.

Denote the first order difference of RRI as dRR where dRR(i) = RR(i+1) - RR(i)

A -1 Hankel matrix of order n is a matrix defined by its first row or column. Given a series A = [a1,a2,a3,.........an], the -1 Hankel matrix defined by A is:

$$H = \begin{vmatrix} a1 & RRa2 & \ldots & an \\ a2 & RRa3 & \ldots & a1 \\ \ldots & \ldots & \ldots & \\ an & RRa1 & \ldots & a(n-1) \end{vmatrix}$$

Such a special Hankel matrix is Hermitian symmetry so it can be diagonalized by a unitary matrix U:

H = U * D U

And the $*$ denotes the conjugate transpose and D is the diagonal matrix with the eigenvalues of H arranged on its diagonal line.We found that the spectrum of H is also central symmetry and the eigenvectors constituting U are bases in Fourier domain. Therefore, we arrange the eigenvalues by ascending order and permute the rows and columns in U*. and U correspondingly. A Discrete Fourier Transform (DFT) is then performed on U and a feature map is obtained (see Figure 4, note that only two non-zeros entries exist for the DFT of each column).

Since the feature map is symmetry, we proposed a descriptor to characterize the structural

irregularity of the original time series A as follows, denote the row index of the non-zero entry in the i th column of U as I(i),i E 1,2,3,............

the proposed descriptor HankDist is defined as :

$$HankDisk = \frac{4}{n^2} \sum_{i=1}^{n2} \mid I(i) - i \mid \tag{3.1}$$

The non-zero entries in the feature maps of pure sinusoidal series lie in the diagonal line and anti diagonal line, so the HankDist for such series will be zero. However, for white noise, the non-zeros entries in the feature map scattered irregularly and the HankDist is large. The HankDist is computed for both RRI and dRR.

(a)

(b)

(c)

(d)
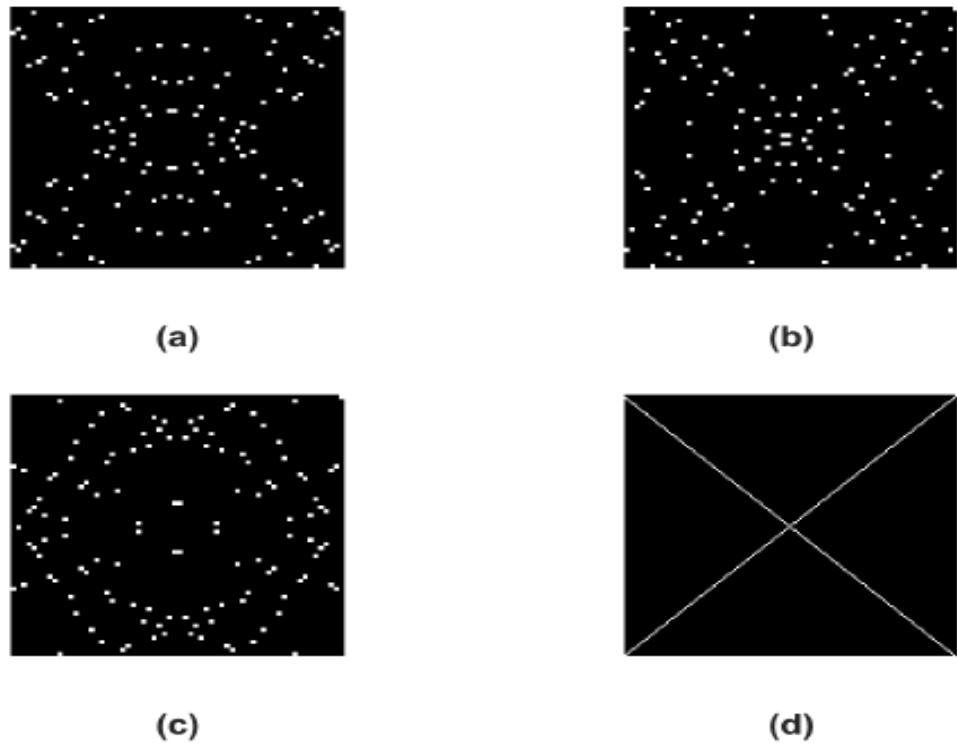
.png

Fig. 3.5: Sample HankDist for: (a) Normal (Record A00003), (b) AF (Record A00005), (c) Other arrhythmia (Record A00008), (d) Sinusoidal signal

Among the standard descriptors [5] about Poincare SD1 and SD2 are linear descriptor of the RRI sequence. Given consecutive RRI pairs (RR(i) , RR(i+1)) are plotted against RRi. In a Poincare plot constructed from normal RRI, most of the data points should distribute in the neighbor of the straight line y = x and the data cloud should form an ellipse. Usually these data points are used to t the ellipse and the long axis (SD1) and short axis (SD2) are believed to reflect the long-term and short-term dispersion of the fluctuation of RRI so also the dynamics of heart activity. Since the length of RRI sequence is very limited, an embedding step is skipped here Denote $(x , y)^T = (RR(i), RR(i+1))^T$, i = 1 , 2 ,3 ,............,N-1,

The data points arerotated 45 degree in clockwise first to extract SD1 and SD2, where the $[x^{'},\ y^{'}\ ]^T$ denotes the coordinates of newdata points. And the variance along x' axes are computed as SD1, and $(SD2)^2 = 2Var(RR) - (SD1)^2$[25].

It has been discovered that although Poincare plot is believed to be of nonlinear analysis tool. Such descriptors (SD1; SD2) are, in fact, linear combination of the basic statistics of the original RRI sequence [25]. Since they can be fully determined by the second order statistics of the original RRI series, they are not nonlinear in their nature. Furthermore, it is obvious that such descriptors are invariant under some permutation s of original RRI series, (e.g. when RR1 = RR(N-1) any shift of original RRI yield exactly the same data cloud on 2-D plane), implying limited ability about reflecting the order structure in RRI.

Complex correlation measure (CCM) is proposed for encoding temporal information of RR sequence and as a real nonlinear descriptor [26]. CCM traces the 'generation' of the Poincare plot by calculating the oriented area spanned by consecutive 3 RRI pairs (after possible delay-embedding). Denote A(i) as the oriented area spanned by,

((RR(i) , RR(i+1)) −(RR(i+1) , RR(i+2)) - (RR(i+2) , RR(i+3))), we have:

$$
A(i) = \begin{vmatrix} RR(i) & RR(i+1) & 1 \\ RR(i+1) & RR(i+2) & 1 \\ RR(i+2) & RR(i+3) & 1 \end{vmatrix}
$$

And CCM is the normalized oriented area:

$$CMM = \frac{1}{K\pi(SD1)(SD2)} \tag{3.2}$$

Different from RRI for which a characteristic power spectrum distribution with interpretable frequency components is expectable [5], the PSD of the difference of RRI sequences (δRR) reveals no significant distinction with respect to that of white noise. However,δ RR was found being capable of distinguish AF from normal rhythm [26]. AFEv is extracted from the Poincare plot of theδ RR and has been validated as a powerful AF detector on baseline datasets including MIT-BIH AF and MIT-BIH NSR [27]. Finally, RR mean and RR std are also calculated.

3.4.2 Spectral features

The HRV based features contain only sketchy information about heart activities. Although the detailed morphology of ECG is vulnerable to noise so it is not a good idea to characterize the CW in time domain, the structure is relatively stable in frequency domain. Welch power spectrum density (PSD) estimation is a modification of periodogram method [28]. Compared with naive periodogram method, reduced variance of the estimation is obtained at the expense of the reduction of frequency resolution.

Suppose we have a uniformly sampled ECG signal of length N. It can be divided into M segments of length L with or without overlapping. Denote overlapping as D, we have:

$$L + (M - 1)(L - D)\text{=}N \tag{3.3}$$

For each segment X(i); i $= 0,1........,$ (L - 1), a windowed discrete fourier transform (DFT) is

calculated as follows:

$$P(n) = 1L \sum_{K=1}^{M} X(i)W(i)e^{-}2ji \tag{3.4}$$

K = 1,2,.........,M;

W(i) is the window function aiming to reduce the frequency leakage.

3.5   Feature selection

Feature selection helps to choose most discriminative features from given feature set and avoid overfitting. A proper balance between the performance on training set and the scale of the whole model improves the robustness of the system and the speed of inference. Since all the features mentioned so far are measured at scales of intervals, a model-free feature selection scheme based on mutual information, called maximum-relevance-min-redundancy (MRMR) is used [29]. This scheme tries to maximize the relevance between the features and the targets (labels) while keeping the redundancies between selected features to a minimal level. Furthermore, for features being measured at the scales of nominal or ordinal, the relevance and redundancy can also be estimated with respect to mutual information.

Denotes the Ω as the set of original features, the S as the set of selected features, t as the targets (labels), I (;) as the mutual information estimation and| . |the cardinality operator, the purpose of feature selection is the find the optimal subset S of cardinality m which:

The straight forward selection procedure to find the subset S is combinatorial, which makes exhaustive search an impractical choice. Furthermore, the estimation of mutual information of high dimensional variable is difficult. In [29], incremental procedure is deployed instead as optimal first-order approximation.

Define the relevance RV and redundancy RD as follows

$$RD = \frac{1}{\mid s \mid} \sum_{i \epsilon s} I(i,j)$$

$$RV = \frac{1}{\mid s^2 \mid} \sum_{i,j \epsilon s} I(i,j)$$

which maximizes RV and minimizes RD simultaneously.

3.6   Classification tasks and additional noise resistance test

Since the proportion of noisy class is very small, 279 recordings belonging to noisy class are excluded for simplicity. The remnant part is a subset consists of samples belong to one of only three different classes.

To investigate the robustness against noise of our method, noises of different intensities are artificially added. Figure 5 gives a sample recording with its noise corrupted versions. Although there might exist noises in original recordings, since we do not have a reliable estimation on the true SNR of them, we take original signals as clean signal. Then we use all the energy of original recordings to compute Ps and white noise of power Pn is added. The signal to noise ratio (SNR) is:

$$SNR = 10log\frac{Ps}{Pn}d$$

For feature values obtained under each noise level, support vector machine (SVM) and bagging trees (BT) are trained as classifiers. Since private test set is not publicly available until now, high entropic k-fold cross validation is firstly used to test the performance of classifiers on all the available data. Then 15% data of each subset (N,A,O) are reserve and only 85%

of data was involved in training and validation, the performances on the remaining 15% data are reported to demonstrate the generalization ability. All the computations are performed on Matlab2015b for 10 times and averaged values are reported.

3.7 Support vector machine

This is a discriminative classifier formally defined by a separating hyperplane that divides data belonging to different classes. The aim is to identify the class of belonging of the different data. Training a support vector machine requires the solution of a very large quadratic programming optimization problem. To resolve this problem the sequential minimum optimization (SMO) technique is used, which is able to divide the optimization problem into a series of smaller possible problems [48]. The classification accuracy can be improve by selecting opportune form and parameters characteristic of the kernel function K(x,y). The most popular kernel function forms are polynomial and radial basis ones. In machine learning, support vector machines (SVMs, also support vector networks are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyperplane that has

the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Training a support vector machine requires the solution of a very large quadratic programming optimization problem. To resolve this problem the sequential minimum optimization (SMO) technique is used, which is able to divide the optimization problem into a series of smaller possible problems. Sequential minimal optimization (SMO) is an algorithm for solving the quadratic programming (QP) problem that arises during the training of support vector machines. SMO is widely used for training support vector machines and is implemented by the popular LIBSVM tool. The publication of the SMO algorithm in 1998 has generated a lot of excitement in the SVM community, as previously available methods for SVM training were much more complex and required expensive third-party QP solvers.

SMO is an iterative algorithm for solving the optimization problem described above. SMO breaks this problem into a series of smallest possible sub-problems, which are then solved analytically. Because of the linear equality constraint involving the Lagrange multipliers , the smallest possible problem involves two such multipliers.and this reduced problem can be solved analytically: one needs to find a minimum of a one-dimensional quadratic function. k is the negative of the sum over the rest of terms in the equality constraint, which is fixed in each iteration.

The classification accuracy can be improve by selecting opportune form and parameters characteristic of the kernel function K(x,y). The most popular kernel function forms are polynomial and radial basis one.

In machine learning, kernel methods are a class of algorithms for pattern analysis, whose best known member is the support vector machine (SVM). The general task of pattern analysis is to find and study general types of relations (for example clusters, rankings, principal components, correlations, classifications) in datasets. In its simplest form, the kernel trick means transforming data into another dimension that has a clear dividing margin between classes of data. For many algorithms that solve these tasks, the data in raw representation have to be explicitly transformed into feature vector representations via a user-specified feature map: in contrast, kernel methods require only a user-specified kernel, i.e., a similarity function over

pairs of data points in raw representation.

Kernel methods owe their name to the use of kernel functions, which enable them to operate in a high-dimensional, implicit feature space without ever computing the coordinates of the data in that space, but rather by simply computing the inner products between the images of all pairs of data in the feature space. This operation is often computationally cheaper than the explicit computation of the coordinates. This approach is called the kernel trick. Kernel functions have been introduced for sequence data, graphs, text, images, as well as vectors.

Algorithms capable of operating with kernels include the kernel perceptron, support vector machines (SVM), Gaussian processes, principal components analysis (PCA), canonical correlation analysis, ridge regression, spectral clustering, linear adaptive filters and many others. Any linear model can be turned into a non-linear model by applying the kernel trick to the model: replacing its features (predictors) by a kernel function. Most kernel algorithms are based on convex optimization or eigenproblems and are statistically well-founded. Typically, their statistical properties are analyzed using statistical learning theory (for example, using Rademacher complexity).

The polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.

Intuitively, the polynomial kernel looks not only at the given features of input samples to determine their similarity, but also combinations of these. In the context of regression analysis, such combinations are known as interaction features. The (implicit) feature space of a polynomial kernel is equivalent to that of polynomial regression, but without the combinatorial blowup in the number of parameters to be learned. When the input features are binary-valued (booleans), then the features correspond to logical conjunctions of input features.

3.8   Descriptive analysis

After preprocessing a 50 dimension feature vector was obtained for each recording, including the HRV based features mentioned above and the power spectrum density estimated

from 0 Hz to 80 Hz.

Rank	SNR=∞ dB	SNR = 30 dB	SNR = 20 dB	SNR = 10 dB
1	AFEv	AFEv	AFEv	RR_{mean}
2	CCM	CCM	CCM	CCM
3	SD_{ratio}	SD_{ratio}	SD_{ratio}	SD_{ratio}
4	RR_{mean}	RR_{mean}	RR_{mean}	$PSD_{4.6-6.9}$
5	$PSD_{4.6-6.9}$	$PSD_{4.6-6.9}$	$PSD_{4.6-6.9}$	$PSD_{6.9-9.2}$
6	$PSD_{6.9-9.2}$	$PSD_{6.9-9.2}$	$PSD_{6.9-9.2}$	$PSD_{9.2-11.5}$
7	$Hank_{RR}$	$Hank_{RR}$	$Hank_{RR}$	$PSD_{2.3-4.6}$
8	$PSD_{9.2-11.5}$	$PSD_{9.2-11.5}$	$PSD_{9.2-11.5}$	$PSD_{11.5-13.8}$
9	$PSD_{11.5-13.8}$	$PSD_{11.5-13.8}$	$PSD_{11.5-13.8}$	AFEv
10	$PSD_{13.8-16.1}$	$PSD_{13.8-16.1}$	$PSD_{13.8-16.1}$	$PSD_{13.8-16.1}$

.png

Fig. 3.6: Top 10 features selected by MRMR

figure 3.1 provides the details about the top 10 features associated with each noise level. It was observed that the linear and nonlinear descriptors based on HRV analysis are almost the top ranked features under different noise levels. When SNR = 10 dB, the noise might disturb the RRI and lead to a different ranking. Generally, HRV based features and PSD from 5 Hz to 15 Hz are those with higher relevance with class labels and bring less internal redundance than other features.

3.9   Classification performance

For two-class classification problems, Sensitivity, Specicity,Accuracy and AUC are well-defined and reported.

For three-class classification problem, we defined the A as positive class and N and O as negative classes and counterparts of Sen and Spec are defined as the quotient of true positive/ negative examples and all positive/negative examples and reported.The results obtained on ar-

tificial test set after 10-fold cross validation was performed on 85% of data during training and validation procedure.
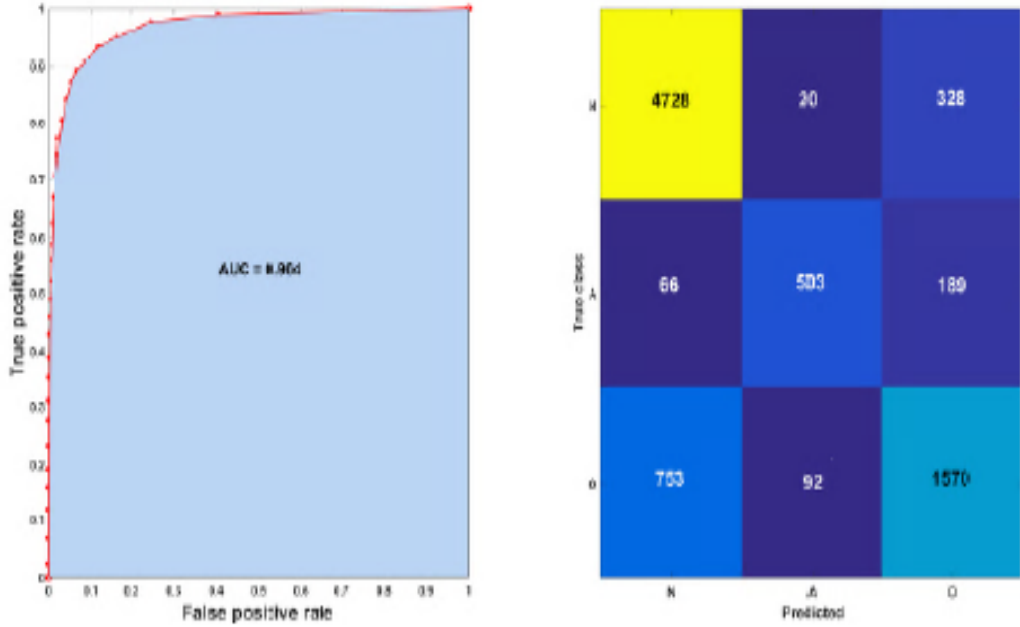
.png

Fig. 3.7: ROC curve and confusion matrix for best performance case

The best performance is obtained for $SNR=\infty$ and all features involved. Figure 3.2 provided the confusion matrix and ROC curve for this case.

3.10   Discussion

The performance metrics degrade slightly across almost all classification problems under different conditions (noise level and classifier configuration) when we eliminate up to 90% of the low-ranked features, which validates the effectiveness of the mutual information based, model-free feature selection scheme. Reduced computational cost and inference time can thus be anticipated in order to adapt to wearable scenarios. Generally speaking, no significant per-

formance degradation was observed when we test our classifier on a subset which was not involved in the training procedure. The generalization ability of our method was validated.

For two-class classification problems, we have observed that very high accuracy 96.6% is obtained when only normal (N) and AF (A) recordings are involved, due to the sensitivity of nonlinear descriptors to arrhythmias. Under 10 dB configuration, the accuracies degrade slightly, maintaining a high level from 92.4% (using only top 5 features) to 95.0% (using all features). However, we have also observed that for other two class classification problems (N vs O and O vs A) the accuracy is off about 10%. Since we do not know the exact condition of patients, the internal high heterogeneity of other arrhythmia class (O) limited the performance especially the sensitivity in N vs O when O is taken as positive class.

For three-class classification problem, average accuracies up to 82.1% are obtained with high specificities. From the confusion matrix for the best case in 7 we found that a considerable part of the O class recordings are misclassified to be N class. It implies that some kinds of arrhythmias behave like normal rhythms with respect to the dynamics characterized by the descriptors we deploy. Since no more details within the O class are available, more targeted feature engineering is difficult.

Moreover, AFEv has reached a nearly perfect performance on benchmark dataset including MIT-BIH AF [27] without much room for further improvement. However, an improvement of its performance on new dataset like CinC 2017 is still possible (94.6% accuracy and 73% sensitivity are obtained under noiseless condition when using AFEv as the only predictor for N vs A task, and only 65% accuracy for N vs A vs O, not reported in Results). Since the benchmark datasets may be still 'small' and 'very clean', cannot cover the exact distribution of data encountered in real-world, hybrid feature may help to improve the robustness of detection scheme, but without the progress from medicine side or big enough data, the questioning about the generality of such machine learning algorithms will never cease.

The involvement of O class has raised interesting questions. For a wide population, we might not have enough prior knowledge about the distribution of certain diseases, neither the medical history of patients. AF and other arrhythmias are all likely to happen. Previous studies on AF detection reported the performance metrics obtained from datasets comprised of

only normal recordings and AF recordings, usually the signal quality is also very good. The prior distribution of benchmark datasets precludes some phenomenon which would always be inevitable if we do not go beyond and give up the pre-assumptions made by the benchmark datasets explicitly, like the existence and thus misclassification of arrhythmias other than AF. The real problem in application scenarios is to detect abnormal ECG recordings and classify them into one of a large number of arrhythmias, instead of detecting AF from normal sinus rhythm. Whats more, when it comes to the application in wearable scenarios, how can we decide to use the detection algorithms of one disease instead of another, or should we run all the algorithms customized for each specific disease and condition in parallel? At least we should take the internal differences between benchmark datasets and real-world data streaming into consideration.

However, we have also observed stable high specificity of the proposed algorithm. High specificity is a welcomed characteristic for preliminary screening of diseases because negative examples take an overwhelming portion in real world scenarios, so high specificity could be beneficial to lower the burden to medical system by excluding non-target population with a high degree of precision. And people who get alarms raised from wearable can get more accurate medical diagnosis from hospital. Generally speaking, we have found that even for relatively short and noise-contaminated ECG recordings, nonlinear descriptors of HRV are still seem to be robust and efficient. Since R peak is the most dominant fiduclpoint in the morphology of ECG and the last characteristic wave to be submerged by severe interference, the lower signal to noise ratio (SNR) for the recording as a whole may not necessarily lead to a worse description about the heart activities extracted from RRI. Heart rate variability based features are thus more stable.

CHAPTER 4

conclusion

An automatic atrial fibrillation detection algorithm was proposed aiming for detecting AF from short ECG recordings since high quality long term recording is not always guaranteed in wearable scenarios for diagnosis where the measurement is always exposed to different kinds of noises and artifacts. The algorithm was validated on a recently released dataset comprised of short ECG recordings acquired by non-standard portable device in non-structured environments. Good accuracy and very high specificity of our algorithm made it a competitive algorithm prototype for preliminary screening. Our research supported the effectiveness of heart rate variability based nonlinear features in AF detection from short ECG recording. Considering the existence of arrhythmias other than AF, cascading classification can be a direction of future work.

REFERENCES

[1] Y.-K. Iwasaki, K. Nishida, T. Kato, and S. Nattel, "Atrial brillation pathophysiology: Implications for management," Circulation, vol. 124, no. 20, pp. 22642274, Nov. 2011.

[2] N. S. Peters, R. J. Schilling, P. Kanagaratnam, and V. Markides, "Atrial brillation: Strategies to control, combat, and cure," Lancet, vol. 359, no. 9306, pp. 593603, Feb. 2002.

[3] S. K. Sahoo, W. Lu, S. D. Teddy, D. Kim, and M. Feng, "Detection of Atrial brillation from non-episodic ECG data: A review of methods," in Proc. EMBC, Aug./Sep. 2011, pp. 49924995.

[4] H. Cecotti, Spelling with non-invasive braincomputer interfaces Current and future trends, J. Physiol, Paris, vol. 105, no. 1, pp. 106114, 2011.

[5] M. Malik et al., "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," Eur. Heart J., vol. 17, no. 3, pp. 354381, Mar. 1996.

[6] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. Zhang, "Automatic Online detection of atrial brillation based on symbolic dynamics and Shannon entropy," BioMed. Eng. OnLine, vol. 13, no. 1, p. 18, 2014.

[7] K. Tateno and L. Glass, "Automatic detection of atrial brillation using the coefcient of variation and density histograms of RR and 1RR intervals," Med. Biol. Eng. Comput., vol. 39, no. 6, pp. 664671, Nov. 2001.

[8] J. Park, S. Lee, and M. Jeon, "Atrial brillation detection by heart rate variability in Poincare plot," Biomed. Eng. OnLine, vol. 8, no. 1, p. 38, 2009.