UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Enhancing The Diagnostic Quality of ECGs in Mobile Environments

by

Taihai Chen

Thesis for the degree of Doctor of Philosophy

April 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

Doctor of Philosophy

ENHANCING THE DIAGNOSTIC QUALITY OF ECGS IN MOBILE ENVIRONMENTS

by Taihai Chen

As the leading cause of deaths worldwide, Cardiovascular Disease (CVD) has imposed a serious burden onto society. Being *reactive* in approach, the current healthcare infrastructure struggles to address the problem properly. In contrast, a *proactive* approach can offer better disease management by predicting impending episodes. The key to the proactive approach is remote continuous monitoring. Traditional long-term monitoring faces serious challenges in transmission as data must be sent out continuously for further processing, leading to short battery life of the sensor node and defying the very notion of continuous monitoring. However, with intelligent signal processing algorithms, signal analysis may directly take place at the sensor node itself. Hence one does not need to transmit data until abnormality is detected. This in turn may save the energy at the sensing node and therefore preserve the notion of continuous monitoring.

In this thesis, we first investigate the automated feature detection of Electrocardiogram (ECG) fiducial points by proposing two different algorithms based on time-domain morphology and gradient, as well as time-frequency-domain with Discrete Wavelet Transform (DWT) respectively. Secondly, to tolerate the possible misdetection errors from ECG fiducial points detection algorithms, we investigate spectral energy as a feature for normal and abnormal ECG classification in feature calculation, as well as statistical analysis of its variation and classification performance under worst-case misdetection. Our exploration shows that spectral energy mostly manages to tackle misdetection error and shows better classification performance than wave duration-based classification. Thirdly, we explore the possibility of adding more time and frequency domain features for enhancing the classification accuracy. Different levels of improvements in classification performance can be observed with respect to the classification models and number of ECG leads involved. Finally, hardware architecture is proposed to integrate spectral energy calculation and Linear Discriminant Analysis (LDA) classifier for an on-body ECG classifier. Verification and power estimation of the system is carried out and shown to be efficient for on-body ECG normal and abnormal classification.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Taihai Chen , declare that the thesis entitled *Enhancing The Diagnostic Quality of ECGs in Mobile Environments* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: as listed in Section 1.4.

Signed:.....

Date:.....

Acknowledgements

First and foremost, I would like to express my whole-hearted thanks to my parents, Mr. Zhaolin Chen and Mrs. Qiongzhen She. Because of the underdevelopment in China only a generatation before, both of them could not afford any higher education at a university. The biggest achievement of which they are, by far, the proudest has been raising their child and sending him abroad for good education and hopefully a brighter future. I owe them the most in my life.

I would also like to thank my undergraduate supervisors in the University of Electronic Science and Technology of China. They are Prof. Liping Li, Prof. Lu Zeng, and especially Prof. Jianguo Ma. It was the Chinese poem written by him that had lightened up my path to study abroad and pursue my PhD degree. Truly he is a great researcher and will have always been an academic idol to me.

I am greatly indebted to my PhD supervisor Dr. Koushik Maharatna for his continuous technical support and, more importantly, having faith in my work and me as a person. It was him who made the decision to offer financial support to an international student for three years time, which truly meant a lot to me. He is also the most diligent researcher I have ever seen in my life. May the best be with him in his academic career.

Regarding financial support, I would like to express my special thanks to the CHIRON Project under EU ARTEMIS joint undertaking, which funded me through my PhD research. Being involved in this project also allowed me to engage in a European project that attracts experts in different domains from all around Europe. I felt honoured to be part of it.

Along the journey of my PhD, there were many researchers and PhD colleagues who offered their considerate hands at different points in my PhD – academic staff like Dr. Srinandan Dasmahapatra and Prof. Mahesan Niranjan who had kindly spared their busy time for discussion with me in Machine Learning; post-doctoral research fellows like Dr. Amit Acharyya, Dr. Evangelos Mazomenos and Dr. Saptarshi Das who had guided me and discussed with me in the beginning and middle phases of my PhD; colleagues like Dwaipayan Biswas, Shre Chatterjee, Wasifa Jamal, Valentina Bono, Sanmitra Ghosh who had kindly shared their experiences in academics and personal lives with me. I am grateful for what they have done.

It is also my pleasure to acknowledge Dr. James Rosengarten, Dr. Nick Curzen and Dr. John Morgan from Southampton General Hospital; Sofia Dima and Christos Panagiotou from University of Patras; Grazia Cappiello from Università di Bologna; Giovanni Baldus and Daniele Corda from University of Pisa; Arnab Bhattacharya from the National Institute of Technology, India. They had been awesome to discuss with.

I would like to acknowledge Matthew Loke, Bogdan Opris, Dayi Chen, Ahmed Rahim, Yutong Han, Zhihua Zhang during their MEng and MSc projects.

I would also like to acknowledge my dear colleagues and friends in my research groups (ESD, ESS and CSPC), to name a few – Dr. Jatin Mistry, Dr. Rishad Shafik, Dr. Saqib Khursheed, Jedrzej Kufel, Dr. Mustafa Ali, Dr. Ke Li, Dr. Sheng Yang, Zuo Cao, Yi Li, Teng Jiang, Dr. Alex Wood, Dr. Alex Weddwell, Dr. Davide Zilli, Dr. Leran Wang, Yang Lin, Wei Wang, Ji Qi, Dr. Thomas Redman, Dr. Robert Rudolf, Jon Storey, Kier Dugan, Dr. Rob Mills, Alex Lam, Dr. Chuan Bai, Shenghao Liu, Shaobai Li, Tianyang Zhou, Dr. Massoud Ghahroodi, Jianhao Xiong, Xiaoru Sun, Dr. Tayyaba Azim, Dr. Ke Yuan, Dr. Wei Liu, Dr. Rares Bodnar, Dr. Matthias Boettcher, Nawfal Firas, Meng Tian, Tristan Aubrey-Jones, etc. for their friendly support throughout my PhD journey.

Lastly, thank you to those that have helped me throughout these years and sorry for not being able to recall your lovely names. I wish the best of luck to you in both your personal and professional lives.

Nomenclature

Acc	Raw Classification Accuracy
ASIC	Application Specific Integrated Circuit
CV	Cross Validation
CVD	Cardiovascular Disease
ECG	Electrocardiogram
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FT	Fourier Transform
GEP	Global Extrema Pair
HDL	Hardware Description Language
HFDA	Hybrid Feature Detection Algorithm
k-NN	k Nearest Neighbour
LDA	Linear Discriminant Analysis
ML	Machine Learning
MMM	Modulus-Maxima Method
MMP	Modulus-Maxima Pair
MRA	Multiresolution Analysis
NG	complexity of 2-input NAND gate
PnR	Place and Route
QDA	Quadratic Discriminant Analysis
Sen	Sensitivity
Spe	Specificity
STFT	Short-Time Fourier Transform
SV	Support Vector
SVM	Support Vector Machine
TDMG	Time-Domain Morphology and Gradient Algorithm
TFR	Time-Frequency Representation
TN	True Negative
TP	True Positive

VCP	Variation of Classification Performance
WCE	Worst-Case Error in Misdetection
WD	ECG Wave Duration
WT	Wavelet Transform

To my parents Zhaolin and Qiongzhen

Chapter 1

Introduction

Recent advances in microelectronics, communications and intelligent algorithms have greatly pushed the boundary of body area network, facilitating various technological aspects like sensor miniaturisation, high data fidelity, and effective signal processing, etc. One common application of body area network is mobile health monitoring. On-body sensors deployed within such systems are generally required to be small in order not to interrupt the user's daily activities. Smaller nodes imply smaller batteries, inevitably creating issues in energy management. One of the main issues is related to data transmission. Research studies on evaluating the broadcasting of data to central server on a continuous basis have been carried out for the last few years [1, 2, 3, 4, 5]. Despite this, investigations on its feasibility via patient trials are still onging [6, 7] before it can be widely adopted in clinical practice. Nevertheless, technically remote sensors are expected to capture and transmit the physiological signal continuously to a central node/server, so that data visualisation and sophisticated signal processing algorithms could be invoked for further data processing. Quite often, sensors are expected to capture and transmit the physiological signal to a centralised node/server on a continuous basis[1], so that data visualisation and sophisticated signal processing algorithms could be invoked for further data processing. That means the radio front-end module embedded in the sensor needs to be fully operational to maintain the notion of continuous monitoring. But such module consumes considerably high energy, and thus the sensor may run out of battery very quickly. To overcome this problem, researches have been done on battery technology and energy harvesting as well as sophisticated coding schemes and compression techniques in order to enhance the battery life and alleviate the burden in transmission. Notably, apart from these strategies, potential solutions may also lie in designing sophisticated signal processing algorithms that could *intelligently* process the physiological data in low-power fashion while *predicting* what and when the sensor should transmit data to other devices. In such way, it may significantly reduce or even eliminate the need for continuous data transmission, in turn preserving the battery in

[1]Broadcasting the data to central server on a continuous basis is fairly common nowadays, see [1, 8, 9].

the sensor while maintaining the notion of continuous monitoring. To achieve this, it is important to explore low-complexity signal processing algorithms that can extract useful and robust information from the physiological data, and thereby *intelligently* decide which part of the data is to be sent for further analysis. In this thesis, we attempt to investigate such methods, statistically analyse their robustness, and design and implement a hardware solution for demonstration purposes.

The rest of the chapter is organised as follows. Section 1.1 justifies the motivation behind the entire research work. Section 1.2 highlights the big picture of remote healthcare monitoring systems and our main research focus in this thesis, followed by Section 1.3 briefly outlining the contents of each chapter. Lastly, Section 1.4 presents the list of publications generated from our research.

1.1 Motivation

According to the recent reports from the World Health Organisation (WHO) [10], CVD remains the biggest cause of death and disability in the world, resulting in 30% of the global total of all deaths per year. Within Europe, 49% of total deaths per year is related to CVD [11]. It has also been pointed out that, throughout the world, the prevalence of CVD will continue in the future. By 2030, almost 23.6 million people are predicted to die from CVDs [10]. Under this context, in the backdrop of a prevailing elderly population together with a shortfall of medical professionals and medical infrastructure, healthcare services with effective disease management in CVD are urgently needed. As one of the main strategic priorities set by WHO, reduction of incidence, morbidity and mortality of CVD can be done with cost-effective and equitable healthcare innovations for management of CVD. In particular, key areas of work have been established, including developing standards of care and cost-effective case management for CVD, as well as means of feasible surveillance with the aim of assessing the pattern and trends of major CVDs and risk factors as well as to monitor prevention and control initiatives [12, 13]. However, healthcare expenditure on CVD has been a serious worry and is foreseen to increase. In [14], the total cost of CVD in Europe has been reported to be €169 billion a year. [15] has also conducted a study showing that the estimated cost of CVD is $296 billion in the United States. In light of the urgent situation in CVD throughout the world, advanced healthcare systems with intelligence, efficient processing capability, and sophisticated, cost-effective disease management is desperately required.

Despite the good wills, current healthcare structures cannot provide the strong backbone of such a visionary system. The fundamental problem lies in how current healthcare systems operate – it has been categorised as a *reactive* approach, where the care is only delivered after an event has occurred. By the time the care the patients needs might be finally given, the patient could have long been in mortal danger.

In contrast to the current healthcare system, the trend of the advanced one is gradually emerging, taking a completly different but more promising approach – a *proactive* approach. This kind of approach is expected to be able to achieve effective disease management by predicting one's impending episodes, given the patient has already been diagnosed with a heart condition. But realise it, continuous monitoring is the key, with which integration of sufficiently long-term physiological information of the patient must be available for analysis and proactive prediction. In the previous generation of healthcare, continuous monitoring might only be possible in-bed. With the advent of ambulatory monitoring (e.g. Holter monitoring [16]), patients can undertake daily activity without impairing much of the monitoring. Despite this, battery life, comfort, and robustness among other factors are always an issue with such monitoring. But now, the way of monitoring has changed. As a collaborative effort, CHIRON [17] is a European research project that intends to combine advanced technologies and innovative solutions into an integrated framework for person-centric health management. In this project, an integrated system architecture of patient monitoring has been proposed to realise the concept of a *continuum of care*, where the health management infrastructure could be deployed at home, in the hospital and in a nomadic environment as well, allowing unhindered daily activity of the patient from wired to wireless monitoring. To support such *remote healthcare*, three important technologies have been the driving forces: wireless technologies, cloud computing and ambient intelligence [18]. With wireless technologies the patients can be free of location dependency and connectivity, so care can therefore be given anytime anywhere; second, cloud computing provides ways for users to use the resources in the cloud very easily on any mobile or fixed devices; and third, ambient intelligence is able to offer intelligent sensors under resource-constrained circumstances (e.g. low energy supply, limited computing resources) as well as context-aware applications.

Speaking of ambient intelligence, a wealth information can be monitored via different types of remote sensors, including physiological information (ambulatory blood pressure, glucose, etc.), biokinetic information (acceleration and angular rate of rotation of human movement), and ambient information (humidity, light, sound pressure level, etc.) [19]. Among them, ECGs are one kind of physiological signal of the human body that effectively reflects the clinical condition of the heart. ECGs are also considered to be easily accessible compared to more elaborate ones like medical imaging with Magnetic Resonance Imaging (MRI). Especially in the mobile setting, ECGs are commonly used to obtain heart information apart from photoplethysmography (PPG) [20, 21]. Rather than performing sophisticated assessments, ECGs are only used as the first screening tool even in a clinical setting where standard 12/15 lead ECG devices are available. That means that a preliminary clinical decision is mostly made based on ECGs. Once ECG analysis is done by clinicians, further clinical investigation with elaborate facilities may then be required [22]. It further leads us to believe that the clinical expectation for a remote healthcare system for CVD is to indicate whether there is any heart abnormality, irrespective of the specific condition causing the abnormality. Therefore in the context

of remote CVD monitoring, the main role of ECG is **to classify the normal and abnormal heartbeats and accordingly produce an alarm**. Reasons for classifying heartbeat into these two broad categories instead of detailed ones can be drawn as two folds:

- Any attempt for specific disease diagnosis could be hazardous in a mobile ECG setting, due to the very limited amount of information provided by the scarce number of leads. For example, some ECG morphological changes that are indicative of certain diseases may not be captured as they may only be seen at particular leads, and these leads may not be available in the remote system.
- Various co-morbidities or confounding conditions may manifest themselves in a similar way in ECG signals despite being different clinical conditions. For instance, various heart diseases are known to exhibit similar morphological changes of the QRS complex (e.g. QRS duration) [23]. So simply considering these as features and trying to classify a disease based on them is not practical, as this in principle will result into a one-to-many mapping [22].

Besides the clinical concerns that result into such decision, one of the huge benefits by doing so lies in technical side where, as later we will see in the thesis, classification for these two classes would be easier in algorithm design, data analysis and associated hardware implementation. The idea of having abnormal class against normal class is fairly similar to *novelty detection*, where departures from normal behaviour are classified as novel events (in other words, abnormal events) [24]. However, instead of grouping the rest of cardiovascular diagnostic classes (which is enormous) except normal control into abnormal class, in this thesis we focus on specific diagnostic classes and regard them as abnormal class. These diagnostic classes include acute events like Myocardial Infarction (MI) and cardiomyopathy/heart failure that may lead to ischemia or arrhythmia episodes, and chronic conditions like bundle branch block, myocardial hypertrophy, myocardial scar and so on that may lead to arrhythmia episodes as well. Furthermore, the effect these two types of abnormal class would be expected to have on the morphology of ECG is of huge difference to normal class and also of diversity - the said acute type normally exhibits ST segment deviation, Q wave missing; and the said chronic type normally exhibits longer TP segment, relatively short ST segment, fragmentation in QRS [11, 25, 26] etc.

Having said so, the key to achieve ECG classification of normal and abnormal classes by continuous monitoring under the setting of remote healthcare is truly the deployment of heterogeneous devices in both wired and wireless fashion. With the advent of wireless sensor network (specifically body sensor network for our case [27]), deployment of a number of appropriate sensors on a patient's body to collect vital biomedical data can now be possible. These data can be processed on the sensors/servers, and combined

with the patient's history to confirm the clinical status with backing of practical clinical knowledge. From there, the clinical decision is thus obtained and further used as an alarm indicating the possibility of an impending episode. Once it happens, care can be directly given to the patient in danger. Meanwhile, the collected data can also be transmitted to an appropriate facility for more elaborate clinical analysis. The entire process indicates that effective management of chronic CVD patients can thus be possible. Preventive intervention by the clinicians may be executed even before the symptoms of a critical episode is fully manifested. Overall, this *proactive* rather than the traditional *reactive* approach may not only bring a significant reduction in mortality rate, but also, economically speaking, enable considerable cost-saving by minimising hospital admission rate, bed time and costly human intervention, as well as minimising socio-economic productivity loss.

1.2 Research Focus

The general conceptual architecture of a remote healthcare monitoring system could be divided into three layers [19, 28, 29, 30] – Local sensor network, Communication and Services, as shown in the upper panel of Figure 1.1. In the first layer, an ECG sensor node connected to various electrodes (either via wire or wirelessly) is attached to a patient's body. From there, vital signs of the person under monitoring can be captured. Depending upon the application, these digital data may be processed partially at the sensor node, or transmitted to the master node wirelessly followed by some signal processing tasks being run at that node. Upon possible request, the raw data or outcome of applications may be shown on the master node to provide feedback to the patient. In the second layer, GPRS / 3G / Wifi / Bluetooth can be used to serve the communication purpose. Through this layer, the raw vital sign data as well as other necessary information (e.g. alarm, outcome of the signal processing tasks) can be transmitted seamlessly to the network. These data can then be shared by the stakeholders in the third layer. Here, the healthcare server is taken as the centralised storage data depot and high performance computing facility, where the data can be further processed under highly complex routines for data analysis, data visualisation and decision making. Meanwhile, other members of this layer including emergency service, physicians and care givers also have the access to all available information about the patient, both via means of wired and wireless communications. Overall, through these three layers, much more efficient healthcare services under this ecosystem of remote healthcare can be delivered to all the stakeholders, most importantly patients themselves, in a seamless way.

Although the system may look ideally powerful, there exists a serious problem in transmission. As mentioned previously in Section 1.1, continuous monitoring is the key concept of remote healthcare. Continuous variability analysis of the data offers more abundant and clinically important information to the doctors, which a *snap shot* of the

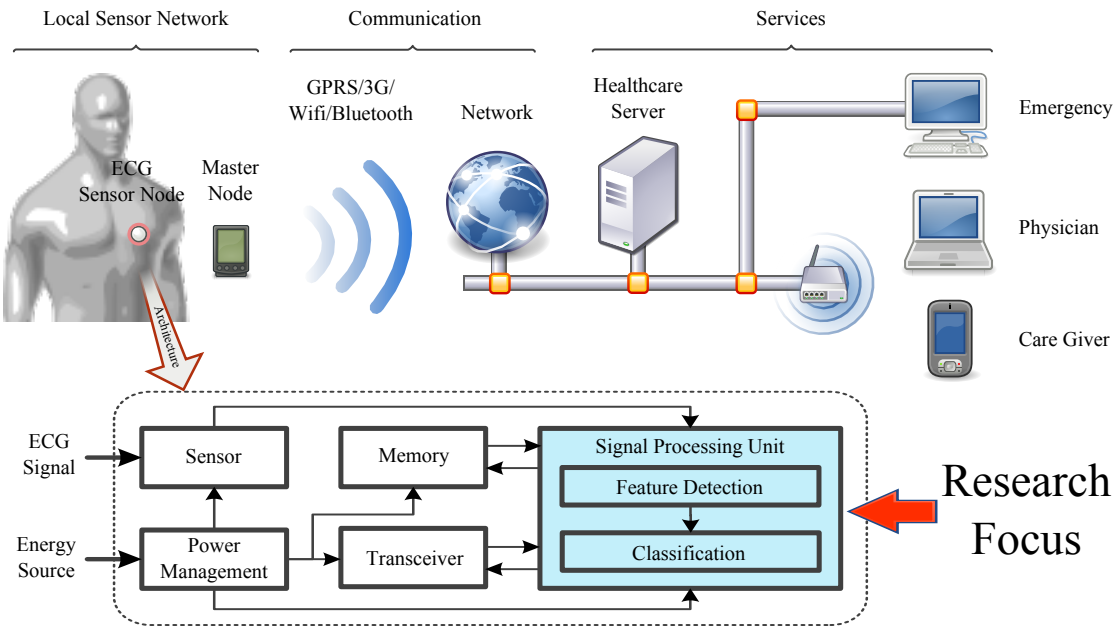


Figure 1.1: Big picture of remote healthcare monitoring system and the conceptual block diagram of the ECG sensor node.

vital sign data cannot provide. To realise continuous monitoring, the traditional way is to transmit the data to a centralised server in a continuous fashion where signal analysis of the data takes place. Since signal analysis is a computationally intensive task, the sensor node itself cannot afford to support such computation-demanding processes due to its limited battery capacity and computational resources. Therefore, most of the tasks are operated on a more powerful centralised server. Though it seems reasonable enough to sustain the notion of continuous monitoring, this traditional approach faces a serious problem in data transmission, particularly from sensor node to master node and master node to network. Continuous data transmission means the most power-hungry radio front-end of a transceiver module within the node is active most of the time, draining the battery very quickly and defying the very purpose of continuous monitoring [1].

To tackle this problem, one possible solution is to come up with an intelligent signal processing unit embedded within the ECG sensor node in the first layer (in the lower panel of Figure 1.1), designed to reduce the usage of the transceiver module. As mentioned earlier, the main purpose of the system is to generate an alarm[2] once abnormality is detected. Therefore, this unit should be designed in such a way that it is capable of processing the ECG data, finding abnormal ECG patterns from the normal ones directly at the sensor node itself, and then generating an alarm in a low-power and yet acceptably accurate manner. That means that unless any abnormality is detected, data transmission would not be required, meaning neither raw data nor an alarm is needed

[2] As later discussed in Sec 5.2 and 5.3, detection of abnormality is not purely based on one single heartbeat. The decision scheme is actually required to analyse multiple heartbeats so as to decide whether or not raising an alarm is needed.

to be transmitted outside. The whole in turn effectively negates the requirement for continuous use of the front-end radio system, saving a significant amount of energy. Furthermore, to maintain the notion of continuous monitoring, the analysed data can be stored in the local memory of the sensor and be transmitted at a preset interval in burst mode upon request [1].

In order to further justify the clinical need for such an engineering solution, it is necessary to provide a brief of the current state of patient monitoring in a clinical community. In general, heart failure is recognised as one of the three long-term conditions together with chronic obstructive pulmonary disease and diabetes [31]. Patients at potential risks for heart failure are normally monitored with ECG in order to keep track of their vital signs and avoid serious cardiac events like arrhythmia, ST-segment ischemia and proarrhythmia based on QT Interval [32]. That being said, whether or not an ECG device/sensor should be used for the patient depends on his/her condition at the time of being diagnosed by the doctor. For cases that normally happen in outpatient clinics, patients are generally not in danger of serious cardiac events, and thus the first diagnosis of the patient is performed with short-term use of an ECG device to observe the ECG and to make a decision. Depending on the diagnosis outcome, the patient may or may not be transferred to an inpatient setting for further investigation, where long-term ECG monitoring may be required. For cases in which admissions are made directly to the emergency department in hospitals, or even an intensive care unit (ICU), these patients are generally at very high risk. Therefore, after clinical diagnosis or even invasive surgery, patients would stay in an ICU or be transferred to a telemetry unit to continue to be monitored, evaluated and treated. This then requires the patients to wear ECG sensors for a prolonged period of time [32].

Recent studies have demonstrated the need of ECG sensors for long-term patient monitoring: a patient trial in [33] aimed to compare the short- and long-term clinical effects of atrial synchronous pre-excitation of ventricles, with or without Cardiac Resynchronisation Therapy (CRT) using remote ECG monitoring and other clinical measures. The outcome of the study was that, by using long-term monitoring, the ECG data analysis indeed proved that there is a long-term improvement in the clinical symptoms of patients with heart failure after the CRT. The differences between optimised biventricular and univentricular therapy appeared to be small for short-term treatment. Another patient trial in [34] analysed the possible improvement of deploying intracoronary Bone Marrow Cell (BMC) therapy to patients with heart failure. Long-term ECG was monitored along with other vital signs at specific times over 3 months to 5 years after intracoronary BMC therapy. The outcome of the study stated that improvement of ventricular performance, quality of life and survival in patients was observed for those who had the therapy. A scalable context-aware cardiac monitoring framework was given in [35]. By providing remote long-term monitoring facilities based on useful ECG analysis, as well as contexts and activities, the proposed system was expected to detect many cardiac arrhythmias. A

prototype was described as proof of concept for the model in this study as well. A long-term wearable ECG monitoring device was also proposed in [36]. This wearable sensor node monitors the patient's ECG and motion signal in an unobstructive way, that the patient's daily life will not be affected while performing cardiac arrhythmia classification in real-time. The outcome of these two studies stressed on the benefits of using remote long-term monitoring for a patient's daily activities. A Wearable Wireless Body/Personal Area Network (WWBAN) was proposed in [37], to intelligently monitor the heart for those with chronic diseases. The outcome of this engineering study clearly stated that remote long-term ECG monitoring for patients was needed, particularly at home to reduce admissions to hospital; however, most of the implemented sensor networks for medical applications are still at prototyping level. New requirements on research concerns in this field, such as energy consumption, have to be taken into account, in order to prolong the lifetime of long-term monitoring. Overall, long-term ECG monitoring for patients with long-term conditions have been justified. Clinical need for such systems is clear in clinical practice, thereby long-term cardiac activities can be analysed and based on which solid clinical conclusions can then be possible.

In terms of patient monitoring with ECGs, there are essentially two types of patients who need ECG monitoring: (1) type I who need monitoring after serious (quite often acute) cardiac events like resuscitation from cardiac arrest or postacute myocardial infarction (MI); (2) type II who have been discharged out of hospital and are at a low risk of near-future cardiac event. For type I patients, ECG devices are necessary because they are in most cases monitored for evaluation and prompt treatment for short periods of time, normally within 24 - 48 hours according to [32]. For example, a remote ECG device is deployed when arrhythmias are thought to be causative in patients with transient symptoms with continuous recorders in hospital settings. Both studies of [38, 39] have stressed on this point, while stating the diagnostic advantage of more comprehensive and real-time data from remote ECG monitoring in hospitals. Apart from that, a remote ECG device is also needed when monitoring is required over a 24-hour period before hospital discharge of myocardial infarction (MI) survivors. [40] stated the clinical need of such, particularly when assessing the risk in patients without symptoms of arrhythmias after MI. On the other hand, type II patients who are discharged for one or two weeks in duration are actually at a low risk for a cardiac event. Both studies of [41, 42] have recognised the need of tele-home-care for these types of patients, and associated solutions are proposed to facilitate the monitoring.

As one of the commonly used remote ECG devices, the Holter device [16] does not require real-time transmission at all. It only needs to store the data (normally in intermittent fashion), and then transfer the data to care givers when the next admission to the hospital is made. In this case, battery charging is not quite an issue, as no potential high-risk cardiac event would be expected and therefore out-of-battery for the device would not cause a serious problem to the patients. Despite this, our engineering solution

facilitate real-time measures in both type I and type II patients: type I patients in the telemetry unit who wear remote ECG sensors for a certain duration are at high risk for cardiac events and close monitoring is required at all times. This low-power solution can then fit in, capturing the ECG data and processing it in real-time, while data is only sent out wirelessly to the central server for processing or display purposes should abnormality be detected. In this way, frequency of battery charge-up is further reduced, and more importantly no potentially significant heartbeats would be missed under longer battery lifetime. For type II patients who are not keen on charging up quite frequently, only abnormal beats are stored in the memory of the device. Should the care givers require, the device then sends out the pre-stored data wirelessly to them for further analysis.

The needs of patient monitoring for prolonged periods of time indeed motivates the clinical need for the engineering solution proposed in this thesis. Despite the good will, to realise our vision above, several elements of the signal processing unit must be investigated, implemented and validated:

- detecting/extracting the clinical features of the ECG signal;
- exploring robust features (i.e. biomarkers) for normal and abnormal ECG classification;
- investigating the possibility of low-power solution to ECG classification from classification algorithm perspective;
- designing and implementing an Application Specific Integrated Circuit (ASIC) architecture for our classification purpose.

Since the signal processing unit will be running on a resource-constrained sensor node, the keys of our investigations are low-power, robust and accurate. In terms of low-power requirements, the algorithms involved in processing must be of computationally low complexity; for robustness, the prospective feature used for classification must be consistent against any artefacts; and as for accuracy, the output of ECG feature detection and ECG classification must be accurate and clinically acceptable[3]. With these in mind, we expect to achieve an efficient and effective solution for ECG classification in a systematic way.

1.3 Thesis Outlines

The rest of the thesis has been structured as follows, with brief diagram given in Figure 1.2:

[3] As will be discussed later in Sec 3.4, CSE standard is regarded as clinical standard for ECG feature detection; for ECG classification, specific validation processes for devices with concern of Major Level by Food and Drug Administration (FDA) [43, 44] have to be investigated, which is out of the scope of this thesis and thus not covered.

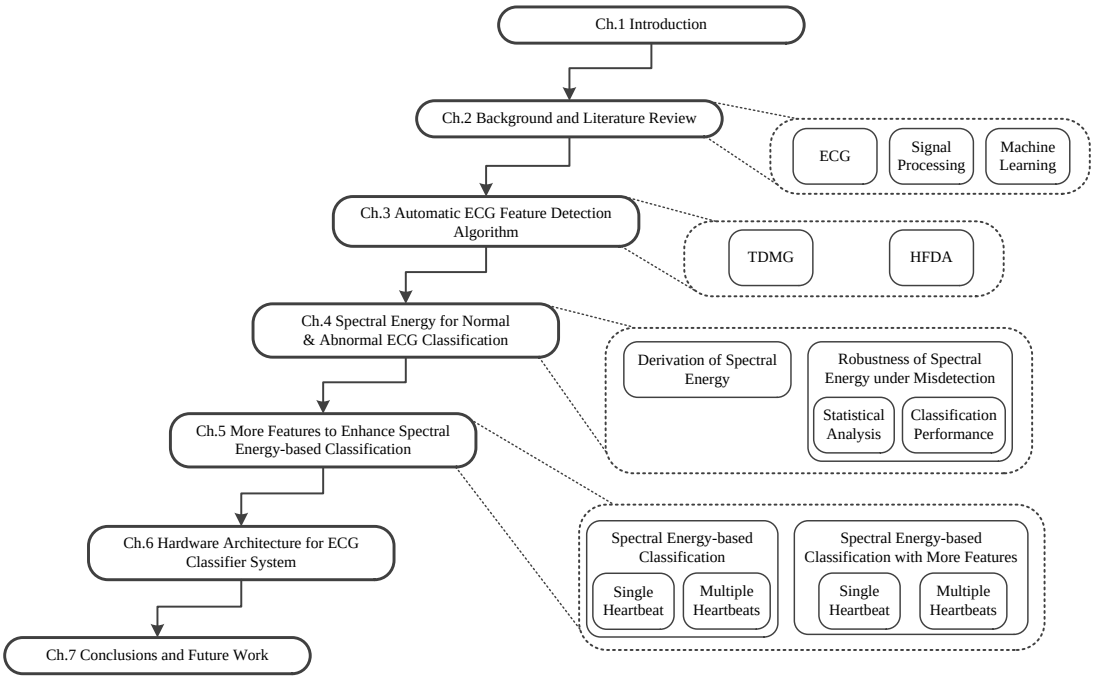


Figure 1.2: The structure of the thesis.

Chapter 2 presents the detailed background and literature review of the ECG, signal processing and machine learning. Firstly, the basic operations of the heart – our main research object, the ECG – are covered, including the basic anatomy, electrophysiology and electrocardiography, and the standard as well as other commonly used cardiac monitoring lead systems. Secondly, discussion on digital signal processing in terms of history, influence, and application is given. Following that, preliminaries of wavelet transform is presented, including basic theories and comparison with Fourier transform and short-time Fourier transform. Also, a number of technical concerns about digital signal processing specifically for ECGs are discussed. Lastly, as a major part of the theoretical fundamentals for our research in this thesis, machine learning is also reviewed. A brief survey of applications in biomedical engineering is discussed, including theoretical background of feature extraction as to feature construction and feature selection, classification algorithms specifically introducing our five selected classifiers used throughout the thesis, and lastly the performance assessment for classification.

Chapter 3 discusses in detail the automated feature detection of ECG fiducial points, the time points at which clinically important cardiac events happen. Since accuracy and computational complexity of the algorithms are of great concern, two different approaches targeting at high accuracy and balance between accuracy and complexity are respectively proposed, implemented and verified against our database. The first algorithm follows a Time-Domain based Morphology and Gradient (TDMG) approach, which includes pre-processing to filter the ECG noise, initial extraction of the fiducial points with predefined searching windows and adaptive threshold policy, and further

refinement with different windows and threshold policy so as to improve the results. The second one takes a different approach, which primarily deploys Discrete Wavelet Transform (DWT) coupled with time-domain morphological analysis to achieve Hybrid Feature Detection Algorithm (HFDA). By using Haar wavelet-based DWT and the resultant DWT coefficient sequences at specific levels, initial extraction and refinement for QRS complex of an ECG heartbeat is done, following which extraction for P and T wave is carried out. Validation of both algorithms in terms of mean and standard division of error against the human-annotated ground truth is made in the end of the chapter.

Chapter 4 details and validates our prospective feature (i.e. biomarkers) – spectral energy for our classification purposes. Since misdetection error is generally inherent in the ECG features generated from the automated detection algorithm, we propose to primarily use spectral energy of specific ECG wave components to serve as the input to classification. After discussing the definition of spectral energy, four different approaches for calculating spectral energy based on the DFT method, the DWT method and thresholding policy upon either the entire ECG or specific wave components are analysed and compared. Each of them is validated upon our five classifiers in terms of classification performance. Knowing that DWT-based spectral energy of specific ECG wave components serves as the best, we further justify the robustness of spectral energy based on this approach by statistically analysing the variation of spectral energy under worst-case misdetection scenario. Similarly, classification performance using spectral energy based on the same approach under misdetection is also discussed. Modified 10-fold cross validation is proposed to implement the experiment.

Chapter 5 gives a detailed discussion on how to enhance the performance of our previous spectral energy-based classification strategy. Based on what we have learned from Chapter 4, certain selections of our spectral energy may not serve well as robust features against misdetection in classification. Light has been shed on encompassing more potentially useful features into consideration and the relevant and non-redundant ones via four feature selection techniques are reasonably chosen, namely ReliefF, InfoGain, CFS and FCBF. In order to demonstrate the usefulness of the enhancement, single and multiple heartbeat classification with our classifiers are experimentally analysed before any deployment of more features, where trade-off along with analysis between classification performance and computational complexity, specifically for Support Vector Machine (SVM), is made. Following which, single and multiple heartbeat classifications with our classifiers are experimentally analysed, but this time with more features to augment spectral energy. Comparison between before and after deployment of more features is also given along the discussion as well.

Chapter 6 covers the flow of designing and validating our ASIC solution to the on-sensor normal and abnormal ECG classification module. For the sake of demonstrating the concept, the whole design follows spectral energy-based classification strategies without the enhancement of additional features. The main part of the chapter presents the

architecture of the system in a systematic way – an overview of the system discussing its key functions and exampled data flow, following which detailed discussion on the design of its sub-blocks are also given. In the end, technical implementation as well as cross-verification between Matlab and post-synthesis of the chip are given.

Chapter 7 summarises the findings and contributions discussed in this thesis. A number of promising research works that might extend and further contribute to our current works are also outlined.

1.4 Thesis Contributions

The contributions of the research work throughout the thesis have been partially published as the following list:

Journals:

- **T. Chen**, E. Mazomenos, K. Maharatna, S. Dasmahapatra, and M. Niranjan, "Design of a Low-Power On-Body ECG Classifier for Remote Cardiovascular Monitoring Systems", *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 3, no. 1, pp. 75-85, Mar. 2013.
- E. B. Mazomenos, D. Biswas, A. Acharyya, **T. Chen**, K. Maharatna, J. Rosengarten, J. Morgan, and N. Curzen, "A Low-Complexity ECG Feature Extraction Algorithm for Mobile Healthcare Applications", *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 2, pp. 459-469, Jan. 2013.
- V. Bono, E. B. Mazomenos, **T. Chen**, J. Rosengarten, A. Acharyya, K. Maharatna, J. Morgan, and N. Curzen, "Development of an Automated Updated Selvester QRS Scoring System Using SWT-Based QRS Fractionation Detection and Classification", *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 1, pp. 193-204, Jan. 2014.

Conferences:

- **T. Chen**, E. Mazomenos, K. Maharatna, S. Dasmahapatra, and M. Niranjan, "On the Trade-Off of Accuracy and Complexity for Classifying Normal and Abnormal ECG in Remote CVD Monitoring System", *IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 37–42, Oct. 2012.
- E. Mazomenos, **T. Chen**, A. Acharyya, A. Bhattacharya, J. Rosengarten, and K. Maharatna, "A Time-Domain Morphology and Gradient based Algorithm for ECG Feature Extraction", *IEEE Int. Conf. Ind. Technol.*, pp. 117-122, Mar. 2012.

The following publication is currently under preparation:

- **T. Chen**, K. Maharatna, "An Investigation into the Robustness of Spectral Energy for ECG Classification in Mobile Environment", to be submitted to *IEEE J. Biomed. Heal. Informatics*.

Chapter 2

Background and Literature Review: Electrocardiogram, Signal Processing, Machine Learning

Necessary theoretical background and relevant application sketch of the three main multidisciplinary fields, i.e. electrocardiogram, signal processing and machine learning, are given in this chapter. It starts with Section 2.1, where the basics of the heart from anatomical and electrophysiological perspectives are introduced. It is then followed by the detailed description of the ECG and its electrical relationship to the heart. Later a coverage of standard hospital-based, derived-based and mobile-based ECG lead system is also given in the hope of justifying the concept of a *lead scenario* used throughout this thesis. Next, Section 2.2 gives a general history as well as the future of signal processing with a focus on digital signal processing. This is followed by the basics of wavelet transform and the comparison of traditional transform techniques. A dedicated discussion on digital signal processing, specifically for the ECG is also given. After that, as one of the key components of our study, a detailed background review on machine learning, including a background in biomedical engineering, feature extraction (feature construction and feature selection), classification models as well as performance evaluation is covered in Section 2.3. Finally, Section 2.4 concludes this chapter.

2.1 The Heart and Electrocardiogram

Throughout our study, the ECG has been the primary research subject. Accordingly, an understanding of the anatomical and physiological basis of the ECG should be placed

at the front. It is very important to prepare ourselves with the electrophysiological fundamentals of the heart and ECG, so as to serve well to our application. Domain knowledge, or in this case clinical knowledge, would certainly facilitate the engineering innovation and development of signal processing and machine learning. So, to start with, Figure 2.1 shows the transition flow of our discussion throughout this section, starting with the heart and electrical condition system as an anatomical category, then action potential, lead system, and finally the ECG. The whole process systematically presents how the ECG is li-nked to each previous block in the following subsections. Note that a more detailed description can be referred to the following materials: [45, 46, 47, 48].

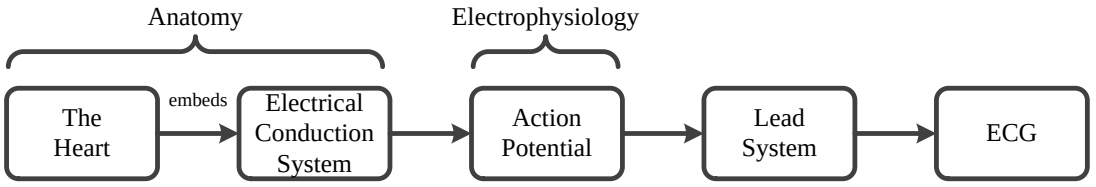


Figure 2.1: The transition diagram starting from the heart to the ECG.

2.1.1 Basic Anatomy

2.1.1.1 The Heart Structure

Figure 2.2[1] shows a cross-sectional view of the heart. It is comprised of muscle called myocardium. As a whole, there are four compartments: the right and left atria, and the right and left ventricles. Usually the heart is oriented so that the anterior aspect refers to the right portion, while posterior aspect to the left portion. In terms of functionality, the most vital task of the heart in general is to maintain the circulatory system of the body, by recycling deoxygenated blood from the rest of the body and then re-oxygenating it so that the heart as well as the rest of the vital organism is maintained. More specifically, it starts with the left atrium retrieving oxygenated blood from the lung via the pulmonary veins, and empties into the left ventricle through mitral valves. From there, the left ventricle injects the blood into the whole body via the aorta. The peripheral circulation system consumes the oxygen and nutrition, and the returns the blood with carbon dioxide as well as waste back to the right atrium via the superior/inferior vena cava. The right atrium then pulls the deoxygenated blood into the right ventricle through the tricuspid valve. After that, the right ventricle forces the blood through the semilunar valve and pulmonary arteries into the lung, where oxygen diffuses into the blood and is exchanged for carbon dioxide. Overall, the heart effectively acts as a blood *pump*, and plays the primary role in maintaining the functionality of this continuously looping circulation system.

[1]This figure is reproduced from Wikimedia.org/Heart_Diagram-en under license CS BY-SA.

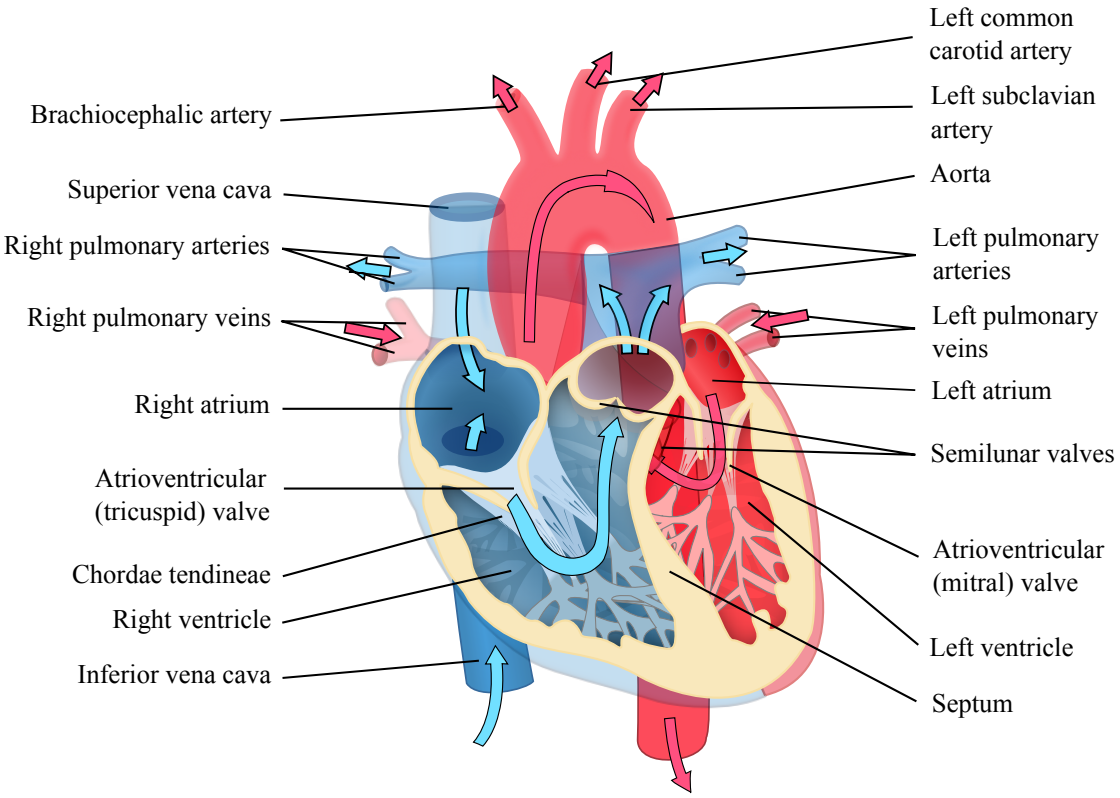


Figure 2.2: Structure of the heart and blood flow running through atria and ventricles.

2.1.1.2 Electrical Conduction System

Behind circulating the blood throughout the body, the heart is in fact rhythmically driven by *forces* to contract. Right before a contraction at a certain part of the heart, a wave of electrical currents passes through and triggers myocardial contraction at this specific part. Accordingly, the entire heart contracts following the path of the electrical current. Here, this path is referred to as the electrical *conduction system* (Figure 2.3[2]). The electrical conduction system of the heart is built up of specialised cells. Among these cells, some of them are specialised in pacemaking and some in the transmission of the electrical current that travels through them. They are dedicated to transmitting the current in an organised fashion to the rest of the myocardium.

As shown in Figure 2.3, located in the right atrium at its junction with the superior vena cava is the *sinoatrial node* (SA node). The SA node is a specialised muscle cell that is self-excitatory and originates the electrical impulse, therefore the electrical current, at an average rate of 70 per minute. Through the three internodal pathways (anterior, middle, and posterior) in the walls of the right atrium and the inter-atrial septum, the electrical current propagates throughout the atria and reaches the *Atrioventricular node* (AV

[2]This figure is reproduced from Wikimedia.org/ConductionSystemOfTheHeartWithoutHeart under license CS BY-SA.

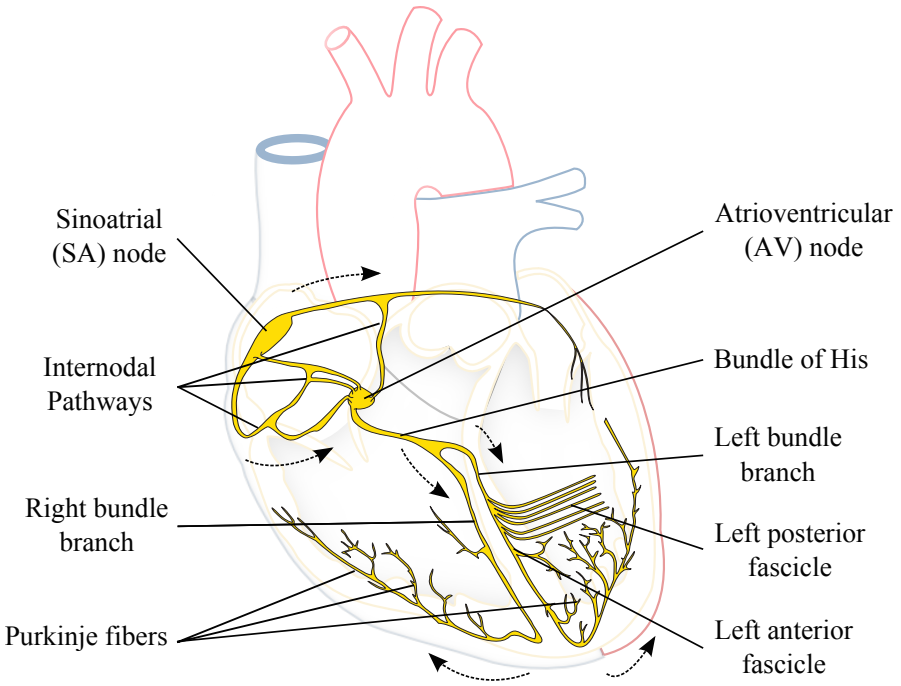


Figure 2.3: The electrical conduction system of the heart.

node). Here, the AV node is located in the wall of the right atrium next to the opening of the coronary sinus and the septal leaflet of the tricuspid valve. Its function is to slow down the conduction from the atria to the ventricles sufficiently so that the atria can contract, fully allowing the ventricles to be adequately filled in order to recirculate the blood at maximum level. Next, *bundle of His* (named after German physician Wilhelm His) follows the AV node and directly leads to left and right bundle branch. From there, the left and right bundle branches travel through each side of the interventricular septum separately. Both branches give rise to fibers that innervate the interventricular septum and the corresponding ventricles. In particular, the left anterior and posterior *fascicles* follow the left bundle branch, where the former one travels through the anterior and superior aspects of left ventricle while the latter one travels through the posterior and inferior aspects of the same ventricle. At the far end, the *Purkinje* cells are present and diverge to the inner sides of the ventricular walls, directly innervating the myocardial cells.

2.1.2 Electrophysiology and Electrocardiography

2.1.2.1 Action Potential

When considering electrical conduction system, the next question is: how is the electrical current created which drives the conduction system and forces the heart to contract?

The answer is *action potential*. In human body, sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}) are the main positively charged ion, and chloride (Cl^-) is the negative one. The way that these charged ions interact with each other by flowing in and out of the cell via the membrane essentially gives rise to the change of action potential, as shown in Figure 2.4 on the left. It then causes the heart to contract.

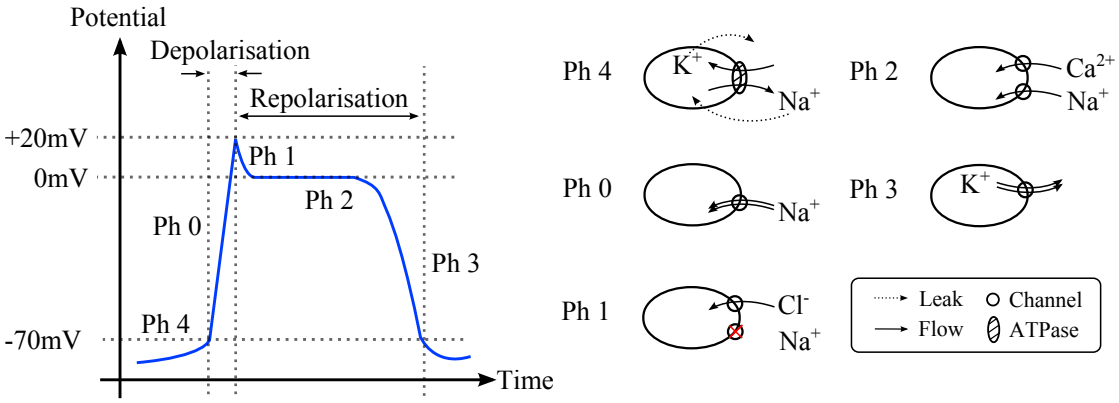


Figure 2.4: Action potential and the corresponding phases of ion movements via cell membrane.

To be specific, let us consider an example. By nature, a live cell tends to maintain differences of charged ion concentrations across the cell membrane. In fact, outside of the cell, there are relatively more sodium and calcium ions, whereas inside there are more potassium ions. Interestingly, the cell membrane is semi-permeable, which means it contains very small leaks that enable in-out exchange of the ions. Thereby, as a starting phase, sodium and potassium tend to leak in and out of the cell (*Phase 4*). However, to maintain the electrical potential, the cell uses pumps (ATP synthase) to move around sodium and potassium ions. As a result, the electrical potential of a resting cell is approximately -70 to -90 mV. But still, due to the leakage effect, electrical potential inside the cell increases slowly and ultimately opens up a new set of channels when it reaches a threshold potential. These channels are dedicated to fast influx of sodium ions. With such big influx of positive ions, a spike of electrical potential occurs soon after the channels open (*Phase 0*). At this specific point, the cell is fully depolarised and a depolarisation phase is achieved, reaching its peak positive charge (*Phase 1*). After that, chloride ions start to enter the cell and slow down the influx of sodium, hence closing the rapid sodium channels. This slightly reduces the potential. Now, two types of channels open: the slow sodium channels and the calcium[3] channels. These two channels maintain the depolarised state of the cell, and therefore a plateau can be observed (*Phase 2*). Next, potassium channels open as a one-way-out hole, causing a rapid decrease of electrical potential inside the cell (*Phase 3*). This essentially manifests the repolarisation phase of the cell. After that, electrical potential restores to the resting phase (*Phase 4*) and the whole process repeats again.

[3]Here, calcium plays a key role in contraction. The more calcium, the longer the contraction can be.

Once the cell is depolarised (but not necessarily repolarised), it needs time to recover before it can fire again. To fire a new electrical impulse, generally the cell takes time called a *total refractory period* (TRP) to restore and normally depolarise again. If there exists a stimuli that is great enough than usual, a not fully-restored cell can be depolarised within a shorter period of time, which is referred as *relative refractory period* (RRP). However, there is an absolute minimum of time for the cell to restore, i.e. the cell cannot be depolarised, and it is known as *absolute refractory period* (ARP).

Having known how action potential is generated from a cell, it then forms the basis of the heart contraction. Anatomically, the heart is comprised of small barrels (cells), which further form long bands by fusing the outsides of every other barrel. These bands again are fused together one-by-one to form sheets. During Phase 2 where calcium activates a clamp[4] and further causes the cell to contract, one of the bands starts to contract leading the ones next to it to contract, making the whole sheet shorten significantly. This forms the action of contraction. When barrels relax, the sheet returns back to the original size. Accordingly, this forms the action of relaxation.

2.1.2.2 Basic Components of the Electrocardiogram

In essence, action potential gives rise to the electrical current, and it is the latter that fundamentally drives heart muscle contraction. The electrical current can therefore be regarded as the manifestation of the heart, or more technically speaking, the electrical activity of the heart. Thus, to detect any abnormality of the heart, it is possible to take advantage of the electrical current for this purpose. Measurements of such can be done from the cellular level or from the body's surface. Between the two[5], electrocardiographic measurement, or *electrocardiography*, is currently one of the most popular screening tools in clinical settings. The measurement is done by attaching electrodes to the surface of the body skin and recording the electrical activity by a device. Such recording is generally referred as electrocardiogram, or ECG.

At this point, a few basic components of the ECG must be introduced. Here, Figure 2.5 shows a typical ECG complex, where clinically important components are shown as well. As the main deflections, there are five waves: P, Q, R, S and T. There are also segments and intervals that represents different cardiac events, such as *TP segment*, *PR segment*, *ST segment* as well as *PR interval*, *QRS interval*, *ST interval* and *QT interval*. In the following, a separate description of each of the components will be given to summarise their characteristics and the corresponding cardiac events.

[4]The *clamp* refers to troponin and tropomyosin complex that is capable of bringing together two ratcheting proteins (actin and myosin) and moving them along each other, which ultimately stimulates the cell to contract.

[5]In fact, there is one more way to measure: without any contact, one may sense electromagnetic activity of the heart through capacitive coupling [49].

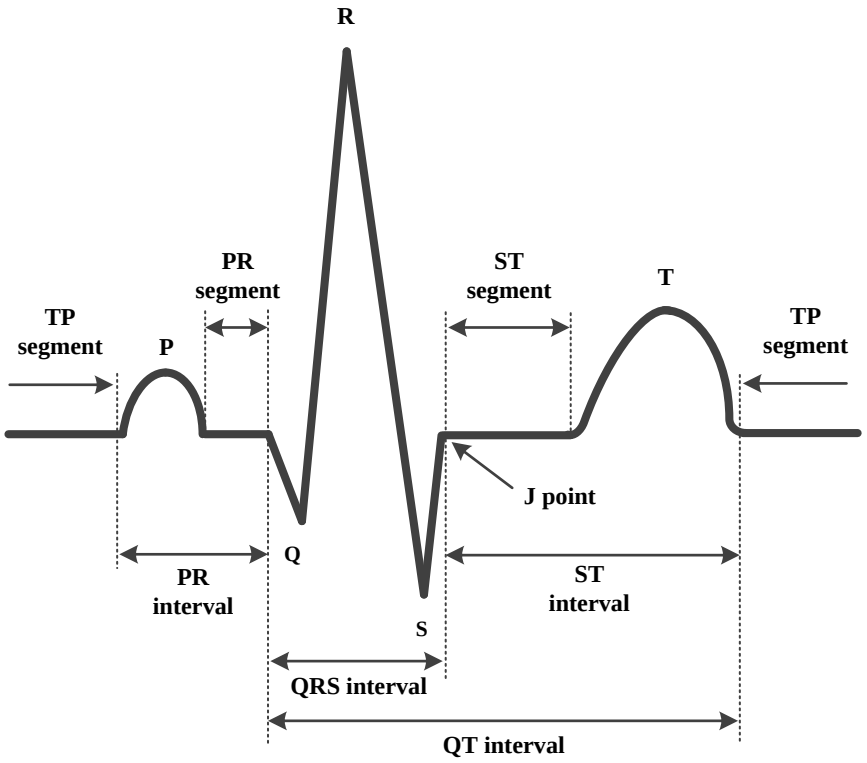


Figure 2.5: Basic components of the ECG complex.

- **Baseline** is an isoelectric line that goes from one TP segment to the next. Ideally it is a flat line throughout the entire ECG complex that acts as a reference for any measurements. However in practice, the baseline would more or less fluctuate due to physiological reasons such as respiration, blood pressure and diurnal rhythms. These would generate low-frequency pattens (usually below 0.5 Hz [50]) on top of the ECG complex. One of the typical examples is Holter ECG monitoring, in which case wandering baseline is very likely to happen [51, 52].

- **P wave** usually comes as the first wave of an ECG complex. It represents depolarisation of both atria, where the first half of the P wave results from the left atrium, and the second half from the right. It also includes the impulse transmission through the three internodal pathways. In fact, the normal duration of the wave can vary between 80 and 110 millisecond (ms).

- **PR segment** is identified right after the end of P wave and up until the beginning of the QRS complex. It represents the spread of depolarisation through the AV node, the bundle of His, and bundle branches, with most of the delay happening at the AV node. This is why it is often found as a flat or isoelectric line.

- **PR interval** essentially represents all the cardiac events from the beginning of the P wave down to the beginning of QRS complex, i.e. the initiation of an electrical

impulse at the SA node up until ventricular depolarisation. The normal duration can vary between 120 and 200 ms.

- **QRS complex** is the major and distinguishable deflection of an ECG complex, as it signifies the depolarisation of the massive myocardium – the ventricles. It usually comprises two or more waves. By convention, the first negative deflection after the P wave is the Q wave, which may or may not be present. Also after the P wave, the first positive deflection is referred to as the R wave. The first negative deflection after the R wave is the S wave. Depending on the altering direction of cardiac vector during ventricular depolarisation, these three waves manifest different characteristics (see Figure 2.7). Normal duration of the complex can vary between 60 to 110 ms.

- **ST segment** is the section from the end of the QRS complex to the beginning of the T wave. The end point of the QRS complex and the start point of the ST segment is referred to as the *J point*. Electrically speaking, the ST segment represents a neutral time for the heart; mechanically speaking, the ST segment stands for the time where the heart is still maintaining the contraction in order to push the blood out of the ventricles.

- **T wave** occurs after the ST segment as either a positive or negative deflection, and essentially represents ventricular repolarisation of the heart.

- **QT interval** covers the QRS complex, the ST segment and the T wave. Essentially it represents all of the cardiac events related to ventricles – from the beginning of depolarisation to the end of repolarisation. The duration of this interval varies according to abnormalities, heart rate, age and sex.

To characterise these components, one way is to locate the most important fiducial points of the wave/segment/interval (e.g. onset, offset, peak). From there, one can quantitatively measure the clinical features of the ECG, which include wave amplitudes, wave durations and interwave durations. Later in Chapter 3, we will show that these clinical features can be extracted to our advantage in our study and further exploited in later Chapter 4 and 5.

2.1.2.3 How the ECG is Constructed

Having shown the action potential and the basic components of the ECG, questions may arise such as "How these two might link together" and electrophysiologically speaking, "How is ECG constructed and ultimately shown as a time sequence". In order to answer these questions, we attempt the following explanation. As is known, there are hundreds of thousands of cells that constitute the heart. More importantly, these cells are lined-up in an organised way to finally build up the electrical conduction system. As a

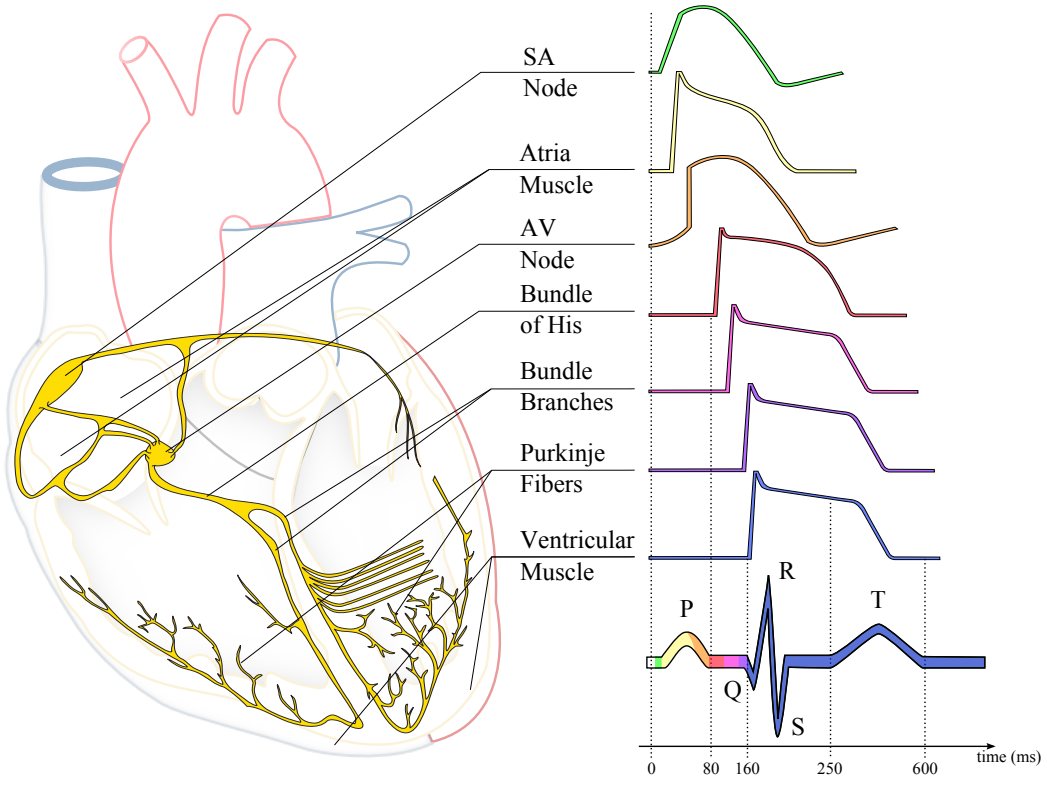


Figure 2.6: The electrophysiology of the heart, showing morphology and timing of the action potential originating from different parts of the heart, and ultimately leading to the form of an ECG heartbeat.

pacemaker cell, the SA node fires an electrical impulse, and it then propagates through the conduction system. Here, the ECG captures the flow of the electrical impulses from the SA node down to the Purkinje fibers. More specifically, the ECG integrates the action potentials generated by cardiac cells through conduction system in time and space during a heart cycle. Figure 2.6[6] depicts the fact that different parts of the conduction system give rise to varied action potentials at different times and different locations. The relationship between these action potentials and the final manifestation as an ECG is clearly shown in colour as well.

Adding to Figure 2.6, Figure 2.7[7] gives a detailed view of instantaneous depolarisation and repolarisation throughout the heart, on top of which the correspondingly resultant instantaneous electrical heart vectors and progressing ECG morphology are also present. Nine different temporal states are shown. Each of them electrophysiologically corresponds to the sequential status of the heart during a heart cycle.

[6]This figure is inspired by Figure 6.7 in [45], and is reproduced from Wikimedia.org/ConductionSystemOfTheHeartWithoutHeart under license CS BY-SA.

[7]This figure is inspired and reproduced from [53].

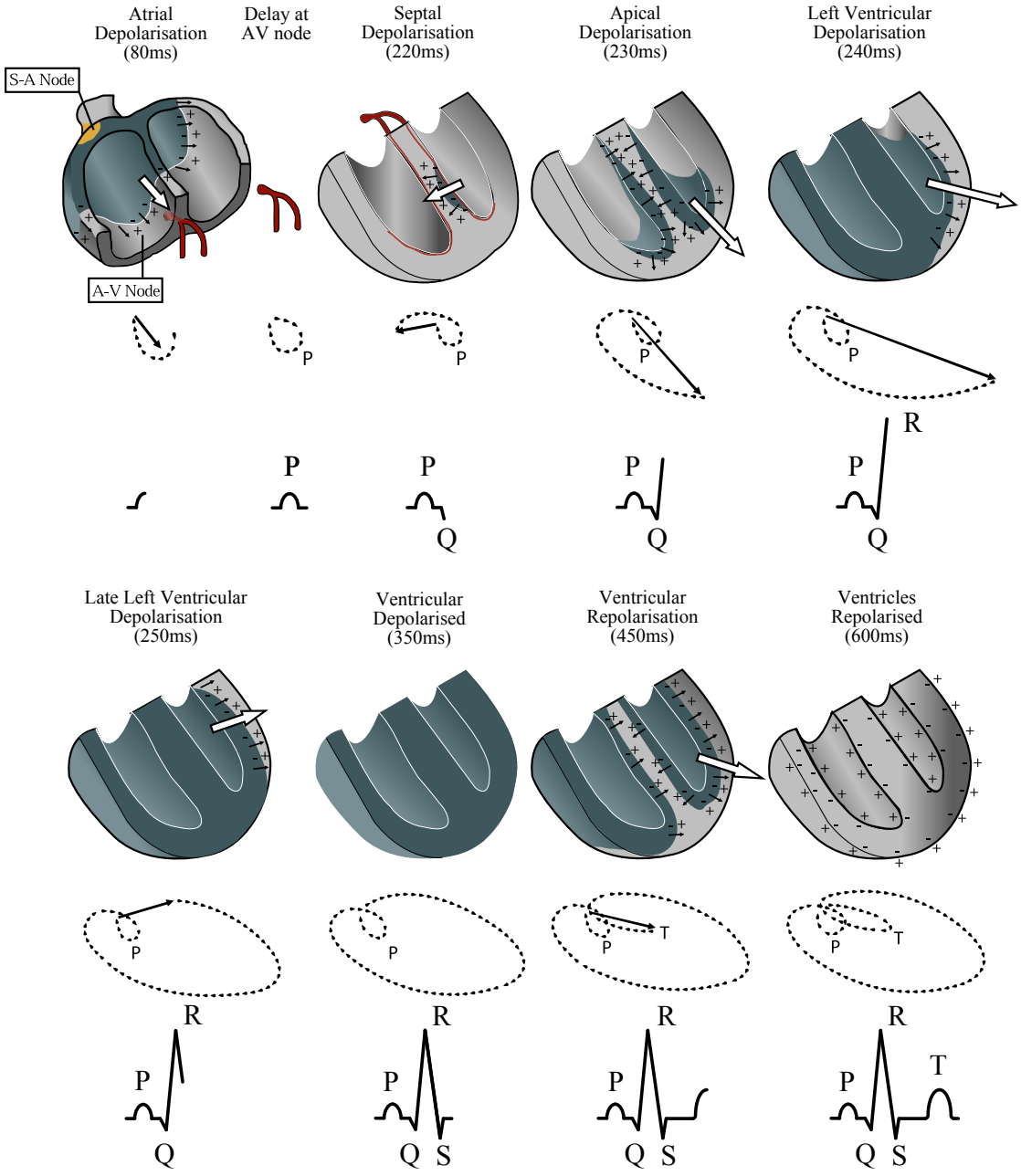


Figure 2.7: The normal sequence of ventricular depolarization and repolarisation with corresponding trajectory of cardiac vector and progressing ECG morphology.

Starting from *atrial depolarisation*, the electric impulses spread from the SA node throughout both atria. The resultant vector is illustrated with a thick arrow and mainly pointing down to the subject's left. Notice that the P wave has started to show up. When the depolarisation reaches the AV node, delay to the activation progress occurs. This delay allows enough time for the ventricles to get filled. By this time, the P wave is formed. Next, the electrical impulses pass through the bundle of His as well as the bundle branches. From there *septal depolarisation* is observed, where the septum starts to depolarise from left to right. This allows the resultant vector points to the subject's right. Soon after septal depolarisation, *apical depolarisation* takes place. Because depolarisation waves occur on both sides of the septum, the wave moving left is balanced out by the wave moving right. Therefore, the resultant vector points to the apex of the heart. Following is the *left ventricular depolarisation*. Since the left ventricle is larger than the right, activation of the left ventricle continues and dominates even after completion of the depolarisation of the right ventricle. In *late ventricular depolarisation*, activation progress of the ventricles is demonstrated even clearer. But since the depolarisation wave propagates along the left ventricular wall toward the back and almost reaches to the end, the magnitude of the resultant vector decreases along the way. After that, both *ventricles* are *depolarised* fully, meaning that there is no activation wave propagation within the ventricles, and hence no measurable cardiac vector. Notice that, so far QRS is formed and hence the depolarisation of the ventricles is fully demonstrated. Following the completion of depolarisation, *ventricular repolarisation* begins from the outer side of the ventricles and further propagates inward. Since both the polarity and the direction of propagation of the repolarisation are opposite to those of depolarisation, it is in fact reasonable to observe that the inward repolarisation wave produces the resultant vector that is analogous to the depolarisation one. This essentially leads to a sign-alike waveform in ECG, which is the T wave. Ultimately, the entire heart beat (PQRST) is observed when *ventricles repolarised*.

2.1.2.4 The Lead System

As mentioned earlier, the ECG is recorded by a device with electrodes attaching to the body skin. The whole set-up is generally called the *lead system*. Though it sounds fairly obvious to deploy a lead system for an ECG observation, in fact it has gone through a long process before it was widely accepted. Interestingly enough, observation of electromotive changes was made by Waller with an Lippmann electrometer over a century ago. As the first initial finding of such, he placed two electrodes on a man's body surface and managed to show that each heart beat was accompanied by an electrical variation [54]. Around a decade later, Einthoven invented the string galvanometer [55]. Since then, electrocardiography had been advancing rapidly. The clinical significance of the ECG had been demonstrated and justified by Einthoven and Thomas Lewis [56], and further improved by Frank Wilson [57], who had introduced the concept of *Wilson's*

central terminal which enabled the use of precordial leads for recordings on the human chest. These advances were the precursors for the standard 12-lead ECG system we normally see to date. Also, there are other well-known lead systems like the Mason-Likar 12-lead ECG system, the vectorcardiographic lead system, and the EASI lead system. All of them will be discussed shortly.

Because the body contains fluids and chemicals that can conduct electricity, the electrical impulses generated from the heart can be transmitted to the body's surface. As a result, placing electrodes on the different parts of the body can help capture and record these electrical impulses. This, in fact, lay down the fundamental of lead system. Note that a *lead* is referred to as an imaginary line in connection with any two electrodes (or, between an electrode and a virtual terminal). This should not be confused with the chemical element "lead". Also, in the clinical community, the term lead is interchangeable with the term *channel*. For more detailed information, readers can refer to [47] and [32].

Last but not least, the aim of this section is to clarify most of the currently available hospital-based and mobile-based lead systems, and their monitoring systems (with digital logger, centralised computing server, etc.) are also covered. More importantly, this section justifies the potential carrier of the lead scenario (a term mainly used in Chapter 4, 5 and 6) in practice.

2.1.3 Cardiac Monitoring Lead Systems

2.1.3.1 Standard 12-Lead System

Figure 2.8[8] illustrates the standard 12-lead system on a human body. In total there are 10 electrodes (red spot in the figure) required to record the standard ECG. Basically 2 types of leads are recorded: bipolar leads that measure the potential difference between 2 electrodes; and unipolar leads that measure the potential at one electrode with regard to a reference point with constant potential.

For bipolar leads, electrodes are placed on the left wrist (LW) and the right wrist (RW), left ankle (LA) and right ankle (RA). Following gives the mathematical equations of the voltage recorded for three limb leads, namely I, II and III.

$$\begin{aligned} I &= E_{LW} - E_{RW} \\ II &= E_{LA} - E_{RW} \\ III &= E_{LA} - E_{LW} \end{aligned} \tag{2.1}$$

These three leads effectively make up the *Einthovens triangle* (Figure 2.8). Note that the combination of these three, $I + III = II$, an equation holds at any instant in the

[8]This figure is reproduced from Wikimedia.org/HumanAnatomyPlanes under license CS BY-SA.

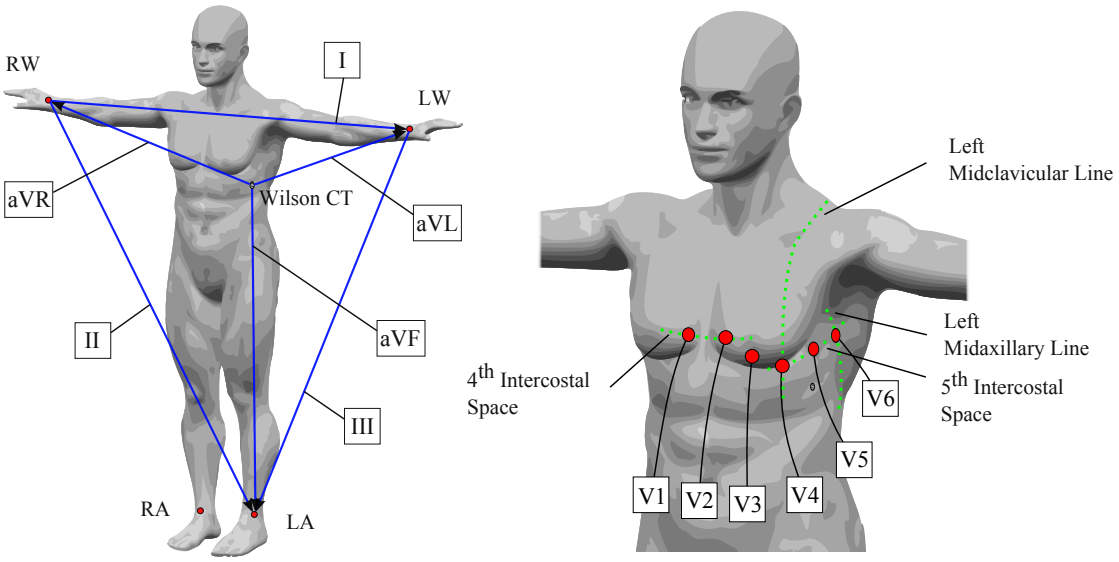


Figure 2.8: Standard 12-lead ECG system. On the left is the limb leads diagram with Wilson's central terminal; on the right is the precordial leads position on the torso.

cardiac cycle, which is known as *Einthovens law*. On the other hand, unipolar leads are divided into two categories: augmented unipolar limb leads and unipolar precordial leads. To achieve these leads, we need a voltage reference point. In fact, summing up the potentials at both arms and the left leg give us a single potential, which remains relatively constant throughout the cardiac cycle. In fact, this summed potential refers as *Wilson's Central Terminal* (CT), which was introduced by Wilson in 1934.

$$E_{CT} = \frac{1}{3}(E_{LW} + E_{RW} + E_{LA}) \tag{2.2}$$

With the use of the central terminal, one can measure the voltage between a specific point P and the terminal, which can be given as

$$V_P = E_P - E_{CT} \tag{2.3}$$

For unipolar lead, though it reflects the potential variation at a single point, technically it works as a bipolar lead because it measures a potential difference between two terminals. If connecting that exploring electrode to, say, the right wrist, we have the unipolar lead of the right arm with respect to central terminal. In order to increase/amplify the voltage measured by the unipolar lead, Goldberger [58] managed to achieve 50% more by removing one of the connections from the central terminal to construct a new terminal E'_{CT} in 1942. The modification was necessary because otherwise the ECG complex would have been too small to observe. From there, he measured the potential difference

between the exploring position and the new terminal. Again, we take the right wrist as an example. Mathematically, we have

$$\begin{aligned} aVR &= E_{RW} - E'_{CT} \\ &= E_{RW} - \frac{1}{3}(E_{LW} + E_{LA}) = \frac{3}{2}(E_{RW} - E_{CT}) \end{aligned} \tag{2.4}$$

Likewise, the same principle can be applied to the other two connections, giving us the following

$$\begin{aligned} aVL &= \frac{3}{2}(E_{LW} - E_{CT}) \\ aVF &= \frac{3}{2}(E_{LA} - E_{CT}) \end{aligned} \tag{2.5}$$

These three modified leads, namely aVR, aVL, aVF are known as *augmented leads* (as they are amplified), and have been standardised as part of the 12-lead system.

Apart from the augmented leads, we have unipolar precordial leads. For this category, electrodes are all placed on the chest (Figure 2.8), where *V1* and *V2* are at the level of 4th intercostal space at the sternal borders, *V4* is in the midclavicular line one interspace lower (i.e. 5th intercostal space), *V6* is at the intersection between the left midaxillary line and the 5th intercostal space. For the rest, *V3* is intermediate to V2 and V4 and *V5* is intermediate to V4 and V6. To derive the voltage, no amplification is needed and simply follow Equation 2.3. Thus far, we have all the leads of the standard 12-lead ECG system covered.

To give a more insightful view of the lead system, a description is made of how the trajectory of the cardiac vector (Figure 2.7) leads to the scalar ECG that we normally see. Figure 2.9[9] depicts a spatial diagram of the 12-lead ECG system at abstract level on the left in conjunction with Figure 2.8. From there the spatial relationship between each participating lead is mapped and linked with respect to the origin. It can be seen clearly but in an ideal form, as the whole assumption is based on not taking into account the heterogeneous property of the human torso [45]. Also, it is possible to see that limb leads lie in the vertical plane while precordial leads lie in the horizontal plane. Imagine the origin is the heart (or, current dipole [60]). The cardiac vector generated in three-dimensional space is projected onto those leads. At each instant of time, each lead reveals the projected magnitude of the vector. To give a clear example, time sequences of one heart beat for 12 leads are shown on the right panel of Figure 2.9. This is specifically tailored to a normal patient. As can be observed, the direction of the lead effectively

[9]The patient subject mentioned can be sourced as patient104 from PTB Diagnostic ECG database as part of PhysioBank, so does the digital data [59].

determines the scalar pattern of the ECG, providing various views of the cardiac vector and thus potential abnormalities. Therefore, synergy between leads can help clinical doctors to make a sensible judgement about a patient.

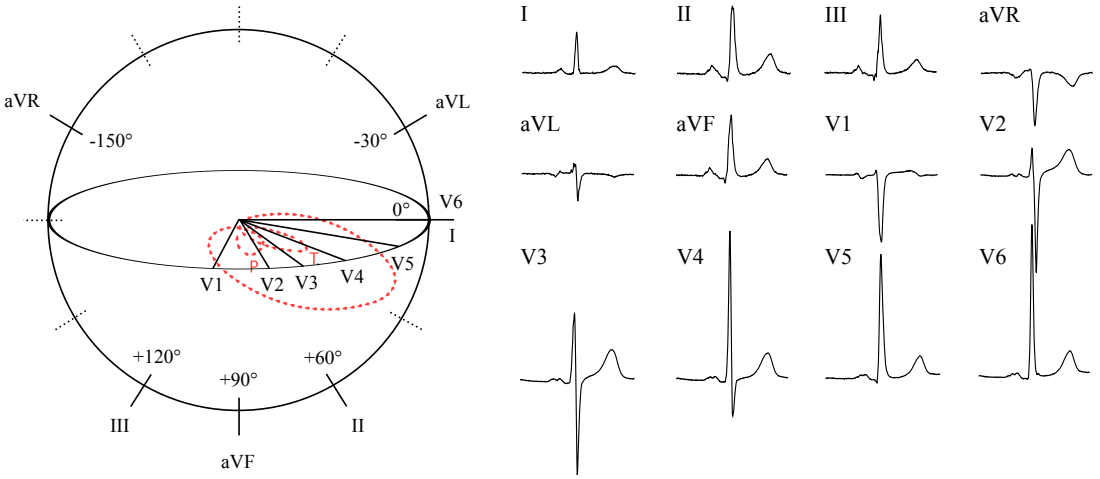


Figure 2.9: An abstract spatial diagram of the 12-lead ECG system, with a panel of time sequences on the right showing one heart beat simultaneously recorded by different leads for a typical normal patient.

2.1.3.2 Bedside Cardiac Monitoring Lead Systems

Apart from the standard 12-lead ECG system, there are other bedside cardiac monitoring lead systems that position the limb electrodes not on wrists and ankles, like the standard 12-lead ECG system does, but on the torso. Doing so helps reduce any artefacts caused by the limb movement as well as avoid tethering the patient, particularly when the patient is exercising or under ambulatory monitoring conditions. Systems that fall into this category are as follows: simple 3 bipolar lead system; limb leads plus 1 precordial lead system; Mason-Likar 12-lead system [61] and vectorcardiographic (VCG) lead system [62].

Obviously, to decide which system to use depends on the practical application scenario as well as the patient's heart condition. For instance, the simple 3 bipolar lead system is particularly useful when performing portable monitor defibrillation, as it is easy to deploy on the go; however, it is not at all appropriate for sophisticated arrhythmia monitoring as it cannot provide a specifically useful lead, in this case V1, for diagnosing arrhythmia. To diagnose arrhythmia, limb leads plus 1 precordial lead (V1) system may be enough, but not for acute myocardial ischemia. In addition, the Mason-Likar 12-lead system is able to provide 12-lead ECG even during exercise stress testing. Though caution should be made as slight differences from the standard 12-lead exists, more importantly, useful precordial leads are available in this system for arrhythmia (V1) and ischemia (V3) diagnosis during patient movement, making it a better candidate than

simple lead systems. Lastly, the VCG lead system is capable of providing 3 corrected orthogonal leads (not geometrically orthogonal but more importantly of equal magnitude and mutually perpendicular): right to left (X), head to foot (Y) and front to back (Z). Based on these leads, clinical parameters that are not possible to observe in a standard 12-lead ECG can be derived and used for diagnosis.

2.1.3.3 Derived 12-Lead ECG based System

Originally attracted in terms of saving time for measurement and storage requirements, the derived 12-lead ECG was first introduced in late 1960s. A matrix of coefficients of linear equations were derived to produce the synthesised 12-lead ECG [63] from XYZ leads as basis leads acquired from the VCG lead system. Though this works as an alternative to standard 12-lead ECG, the value of this method is limited due to the fact that such generalised matrix cannot be accurate for every patient.

Apart from that, reduced lead systems which apply Mason-Likar limb leads plus V1 and V5 (or V2 and V5) were also used as the basis leads in derivation of the 12-lead ECG. With these six standard electrode positions and a matrix of coefficients based on either an individually patient-specific transformation or a population-based transformation, it is possible to perform the derivation for the rest of the leads. The practical advantage of this method is that the originally available leads are secured to be real, not synthesised. Overall, derived 12-lead ECG based systems are believed to retain the clinical potentials, but demand careful practice and comparison with standard 12-lead ECGs.

2.1.3.4 Mobile (Ambulatory) ECG Recorders

Since our main focus in this study is on mobile monitoring, it is very crucial to review what kinds of mobile ECG recorders (or ambulatory ECG recorders, as this phase normally used in clinical community) are available, not only in the research community, but also in commercial market. Unlike standard 12-lead systems and other bedside cardiac monitoring lead systems, originally mobile ECG monitoring was commonly tailored to cardiac rhythm (mainly for the detection of arrhythmias) and transient ST-T changes for possible myocardial ischemia. The most well-known one is the *Holter ECG monitoring* named after its inventor Norman Holter [64]. In general, it comes with a device that records a few channels of ECG data on a patient's chest for 24 to 48 hours, followed later by data processing and physician review. As technology has advanced, with multichannel, digitised and telemetered signals introduced, the Holter ECG has been expanded for more functional purposes, e.g. the analysis of heart-rate variability, QT-dispersion and variability, etc. [47].

Depending on the requirements of the clinical diagnosis as well as the product model, the Holter recorder most commonly uses 1, 2 or 3 leads to acquire ECG signals from 5 or 7

electrodes, namely modified V5 (CM5), modified V3 (CM3) or V2 (CM2), and a modified inferior lead (aVR, reverse Nehb I, Nehb D) [47]. Typical products include the Medilog® AR12 plus from Schiller [65] and the CardioMem® CM 4000 from getemed [66]. The choice of these bipolar leads provides high-amplitude ECG signals and the possibility of detecting the majority of changes in repolarisation (especially for ST-changes).

Furthermore, it is even possible to have 12 leads on the go with more advanced mobile recorders. Contemporary Holter devices, such as the Medilog® FD12 plus from Schiller [65], can offer 12 channels with 10 electrodes, which is analogous to the Mason-Likar 12-lead system but especially targeted at mobile monitoring. Apart from that, an ingenious Holter device was also introduced by Philips called the DigiTrak® XT Holter [67], which is capable of deriving the 12-lead ECG from four (EASI) plus one reference electrodes [68]. As an alternative to the standard 12-lead system, this Holter device offers the advantage of using only five electrode positions over easy-to-locate, bony structures on the torso. Though both devices have shown significant potential and have been extensively researched [69, 70], the Mason-Likar-based and EASI-based Holter devices have been suggested only as a substitute for standard ECGs for monitoring, but not as a substitute for diagnosis [47].

Lastly, the pacemaker is deployed as an *artificial heart pacemaker* capable of correcting an abnormally slow heart rate by taking over the function of the natural pacemaker; similarly the *Implantable Cardioverter Defibrillator* (ICD) is deployed to detect any life-threatening arrhythmias and regulate the heart rate once they occur. Modern pacemakers and ICDs can also be used to gather information about arrhythmias, but mostly with only one lead. For instance, researchers managed to utilise an ICD lead to detect repolarisation alternans (i.e. a beat-to-beat alternation in magnitude of the transmembrane voltage of ventricular cells during repolarisation) that may be useful in improving arrhythmia therapy techniques [71].

2.1.3.5 Discussion

Above all, Table 2.1 shows a list of cardiac monitoring lead systems that sum up the main features of each system discussed above, including the number of electrodes needed to attach on a patient's body, the number of leads (either derived or not derived depending on the systems), and what leads are available for processing and latter analysis. Paying particular attention to these features can be attributed to the following reasons:

1. First, the number of electrodes would certainly affect the comfort of the patient. That means, fewer electrodes would alleviate the patients discomfort, prevent interfering patients daily activities and reduce equipment cost. Technically, fewer electrodes would also mean fewer artefacts when running an analysis, but more due to loose contact, physical movement, etc. if otherwise.

Table 2.1: Cardiac monitoring lead systems.

Method	Title	# of Electrodes	# of (derived) Leads	Leads	Comments
Currently Used Hospital Monitoring Lead System	Standard 12-lead	10	12	Standard limbs and precordial leads	Standard lead system for any particular ECG diagnosis.
	Simple 3 bipolar lead	3	1	I, or II, or III, or MCL_1	Mainly used for portable monitoring defibrillator. To track heart rate, detect R waves, etc.
	Limb leads + 1 precordial lead	5	7	Standard limb leads + 1 precordial lead	Depending on the precordial lead, may be used for arrhythmia or ischemia, but not both at the same time.
	Mason-Likar 12-leads†	10	12	Mason-Likar limbs and precordial leads	Specifically for 12-lead ECG exercise stress testing, and can be used for monitoring arrhythmia and ischemia.
	VCG	8	3	X, Y, Z	For acute coronary syndromes, but not widely used.
Derived 12-lead ECG based System	VCG	8	12	Standard limbs and precordial leads	Derive 12 leads using linear equations based on VCG leads.
	Reduced lead system	6	12	Mason-Likar limbs and precordial leads	Using limb leads plus V1 and V5 (or V2 and V5) to reconstruct the rest.
Mobile ECG Recorder‡	Holter	5	1 or 2	One or two leads out of CM5, CM3 (CM2), aVR	Usually able to record for 24 to 48 hours for the purpose of monitoring arrhythmias and myocardial ischemia.
	Holter	7	3	Three out of CM5, CM3 (CM2)§, aVR	(As Above.)
	Holter	10	12	Mason-Likar limbs and precordial leads	(As Above.)
	EASI	5	12	Standard limbs and precordial leads	Derive 12 leads using linear equations based on EASI leads.
	Pacemaker and ICD	1	1	Apex of atrium or ventricle.	Usually one specific lead goes in either the right atrium or ventricle.

† Recent realisation of such has been made as part of the function of Holter device, as the one with 10 electrodes in the table.
‡ In the clinical domain, an ambulatory ECG recorder is often referred by researchers and engineers. In our study, we generally refer to it as a *mobile ECG recorder*, as 'mobility' is the key term. Note that detection and measurement of rapid/slow heart rhythm, and the diagnosis of causal symptoms (e.g. loss of consciousness) serve as the common uses of ambulatory ECG. For more information on various types and common uses of the ambulatory ECG recorders, the reader can refer to [72].
§ Either CM3 or CM2 is selected, not both at the same time.

2. The number of leads depends heavily on the number of available electrodes and the technical principle of the lead system. Depending on the application scenario, fewer electrodes and fewer leads does not necessarily mean that one system is less functional than the others. With properly sufficient leads, it is acceptable to apply fewer electrodes and fewer leads for designated diagnosis.

3. Construction of participating leads varies between systems and, more importantly, their positions on the body. With the same type of device, the leads that are used may also vary, for instance, Holter ECG recorder in this case.

In clinical settings, depending on the practical environment and the diagnostic application, different lead systems are chosen to address specific needs. In our case, classifying normal and abnormal ECGs in a mobile environment cannot be based on the full standard 12 leads, otherwise the patient would be affected by tethering or discomfort. Instead, a limited number of leads might easily fulfil the task as no specific diagnosis is required (This was already justified in Section 1.1). Speaking of limited number of leads in mobile environment, in fact there are a number of mobile ECG lead systems that do not follow the standard 12/15 lead scheme, for instance most of the wearable sensor-based systems surveyed and reported in [73]. From there it can be seen that, as long as it meets the requirement of medical applications (e.g. general health monitoring, rehabilitation monitoring, etc.), a fewer number of leads is viable. This actually prompts us to think that, we should not limit ourselves into choosing a specific lead system listed in Table 2.1 for our task, but be receptive to a system that utilise various leads from them. In other words, we aim for what we want to achieve, rather than being constrained by the currently available and well-researched lead systems. Therefore, we set up different *lead scenarios* where only a limited number of different leads is available. In this thesis, we envisage there are five lead scenarios, ranging from 1 lead scenario to 5 lead scenario. As their names imply, 1 lead scenario has only one participating lead, whereas 5 lead scenario has five. As we will see in latter chapters (Chapter 4, 5, 6), lead scenarios play an important role in providing the basic practical setting for our classification purposes.

2.1.4 mHealth and its applications in Electrocardiogram

With the dramatic technological advance in today's society, one cannot turn a blind eye to the unprecedented spread of mobile technologies, together with their innovative application to address global/local health problems [74, 75]. Nowadays, there are nearly 5 billion mobile phone subscriptions in the world, where 85% of the world's population are now covered by a commercial wireless signal [76]. Coupling such enormous "mobile population" with healthcare, mobile healthcare comes into the picture. Mobile healthcare (or, mHealth) is a term used for the practice of medicine and public health for health services and information with support from mobile devices [77], such as mobile phones,

patient monitoring devices, personal digital assistants (PDAs), and other wireless devices. On top of these, mHealth utilises not only the mobile phone's core utility such as voice and short messaging service, but also more complex and sophisticated functionalities such as General Packet Radio Service (GPRS), third and fourth generation mobile telecommunications (3G and 4G), Global Positioning System (GPS), Bluetooth, as well as intelligent applications such as those developed under the iOS or Android framework. With these powerful backbone to support healthcare, the motivation behind bringing mHealth to the society is indeed to transform the way health services and information are accessed, delivered and managed, and provides the possibility of greater personalisation and citizen-focused public health and medical care, particularly for those in low and middle-income countries [78].

Despite hundreds of mHealth pilot studies, there is lack of programmatic evidence to inform implementation and scale-up of mHealth, according to the professionals from public health sector [79]. That means, still there is definitely a big gap between mHealth applications and end-user management. To help understanding this gap, [9] analysed the current status of mHealth in disease management, provided an overview of the types of data transmitted, discussed issues of privacy, standards and evaluation. Standing more from an engineer view point, [80] uncovered the great motivation behind the transformation of healthcare, and advocated means of how information technology and engineering can support the transformation of healthcare. These include research areas such as universal language, bioinformatics, semantic networks, data mining, etc.

Among them, sensor monitoring is one of the main focuses in mHealth. Putting sensor monitoring in the context of ECG and associated applications, a fast growing number of engineering solutions have been developed to resolve those most-challenging problems in this field: a role-based intelligent mobile care system with alert mechanism for patients with hypertension and arrhythmia in chronic care environment was proposed in [81], which involves patients, physicians, nurses and healthcare providers. The whole personal mobile device construction comprised mobile healthcare system front-end, physiological parameter extraction devices and mobile phones as personal mobile gateways. low-cost portable real-time cardiac patient monitoring hardware/software co-designed platform with specific focus on medical privacy was proposed in [82]. Similar hierarchical sensor-based healthcare monitoring architecture in wireless heterogeneous networks can also be found in [83]. In 2014, a project called WE-CARE, an intelligent telecardiology system using mobile 7-lead ECG devices, conducted clinical trials at Peking University Hospital. With a clinically acceptable latency around one second, the system was able to achieve over 95% detection rate against common types of anomalies in ECG [84]. To overcome the overload issue of complex ECG data in terms of system integration, a system light-loading technology was proposed in [85] exploiting the manifold-learning-based medical data cleansing to allow mHealth applications in WE-CARE project to achieve higher ECG anomaly recognition rate. To support

low-power mHealth solution at system-on-chip level, state-of-the-art solid-state circuits and systems have been proposed and developed. A very low-power multi-functional ECG signal processor was reported in [86], with several architecture-level power saving techniques such as global cognitive clocking, and circuit-level design techniques such as near-threshold level shifting to achieve 457 nW power consumption at 0.5 V using 180 nm CMOS process. Another mobile healthcare oriented cardiac sensor assisted with machine learning techniques was reported in [87], consisting of modules like cardiac signal acquisition, filtering with versatile feature extractions and classifications. Being able to provide real-time syndrome detections of arrhythmia/vectorcardiogram-based myocardial infarction, the system consumes 48.6/105.2 μW using 90 nm CMOS process, respectively. However, these two solutions may actually encounter serious challenges in real-life deployment, such as patient's discomfort, baseline wandering caused by physical contact between patches and the skin, etc. To overcome these challenges while offering several-years battery lifetime, a small form-factor syringe-injectable ECG recording and analysis device for atrial fibrillation arrhythmia monitoring was proposed in [88], which consumes only 64 nW in 65 nm COME process. This device also offers great advantage from clinical perspective, as it can be injected under the skin near the heart using a syringe needle to avoid surgery, at the same time retaining the benefits of an implantable system. Overall, these state-of-the-art SoC chip designs for mHealth ECG applications truly have great potentials in replacing the current standard mHealth on-body processors to enable higher processing capability and yet more power-efficient solutions.

2.2 Signal Processing

Signal processing is one of the most powerful technologies that execute operations upon signal that represents time-varying physical quantities. It is a technical field that has brought profound changes to other areas, namely space, military, medicine, commercial production, etc. In particular, signal processing has been providing technological foundation for engineering in the biomedical field, especially the processing and analysis of biosignals. This type of signal is produced in the form of energy generated by biological systems, including chemical, mechanical, thermal and electrical energy, to name a few [89]. Biosignals carry corresponding information about the biological systems and thereby reflect the conditions of the systems, just like ECG to heart. Therefore, in order to interact with a biological system, a biosignal is the preferred communication medium. Understanding biosignals would help us to understand the biological system. To achieve this goal, the deployment of transducers can be made to measure the energy. After acquiring the electrical signal from transducers, it is possible to manipulate, process and evaluate the biosignal, thus allowing us to analyse the condition of the biological system.

There are two forms of electric signals that represent our biosignals: the analogue and the digital. In general, these two forms are usually present in a typical bioengineering

system, linked by some key processing elements to get from one to another. As shown in Figure 2.10, a typical bioengineering measurement and processing system consists of six components: first, the transducer is fed by physiological energy which then converts it into an electric signal. Amplification and pre-processing, e.g. filtering, are usually performed. Since the most powerful signal processing algorithms are implemented in digital form, it is often required to have an analogue-to-digital converter (ADC) that converts the signal into a digital format. With this format, the signal can be easily stored in memory, allowing further processing afterwards. Finally, digital signal processing algorithms varying at all sorts of forms and sophistication can then be performed to fulfill the task.

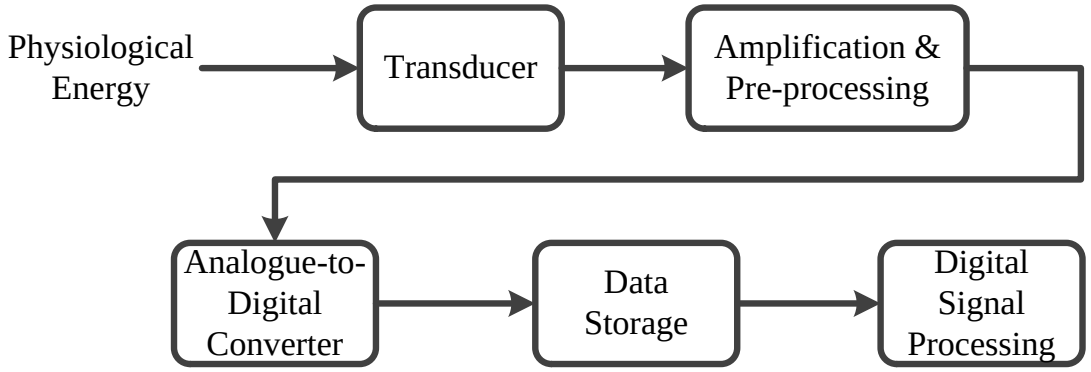


Figure 2.10: A typical bioengineering measurement and processing system.

2.2.1 Digital Signal Processing

Unlike analogue signal processing, digital signal processing (DSP) heavily relies on the digitised data upon which sophisticated DSP algorithms operate. Indeed, DSP is a field of science and engineering that has developed rapidly over the past 50 years. During this period of time, advancement of digital computer technology, new mathematical theories and increasingly sophisticated applications have resulted in the rapid development of DSP. Disciplines like communications, medicine, entertainment, space exploration, etc. have deeply benefited from the technology. The trend of prosper is deemed to continue, and with much more to come. In the following, a brief of DSP will be given from historic perspective, future promise, as well as its advantages and disadvantages.

Firstly, appreciation on the development of this field brings us back in early 1950s, where signal processing was typically done with analogue systems made of electronic components or even mechanical devices [90]. At the time, digital computers were very expensive and limited in their capabilities. They were only used to simulate a signal processing system before implementing associated analogue hardware, simply due to their flexible nature in experimental environment and thus saving economic and engineering resources. Although interesting and sophisticated signal processing algorithms

were developed, they could not be done in real time because of the limited capability of digital computers. As a result, people tended to choose the analogue approach in favour of speed, cost and size. This situation went on until the late 1960s, when significant breakthrough in the methods of efficient computation of *Fourier Transform* (collectively known as *Fast Fourier Transform*, or FFT) was introduced on digital computers [91]. Previous computation time taken as several orders of magnitude greater than real time was significantly reduced, making it possible to run, for instance, spectrum analysis much faster. This also meant that much more sophisticated signal processing algorithms were now practically feasible. The whole turning-point revived DSP as a technique, and made itself an important field of investigation. By the mid-1980s, Integrated Circuit (IC) technology has advanced dramatically [92], allowing very fast and specifically designed fixed-point and floating-point microcomputer architectures for the implementation of DSP algorithms. Memory as well as analogue-to-digital (ADC) and digital-to-analogue (DAC) converters were becoming cheaper and cheaper at the same time. All together were seen to have huge impact on the development of DSP as well. These inexpensive and fast digital circuits have made it possible to construct complex DSP functions, which were generally very difficult for analogue systems. Therefore, signal processing solutions that were previously done in the analogue domain have since been realised in digital domain with cheaper and reliable hardware [90].

Looking at the future of DSP, two aspects should be analysed separately: system-wise and technique-wise. Regarding the former, it is highly anticipated that very complex DSP systems will be implemented with lower cost, smaller size and lower power consumption along with the striving growth of circuit densities and production yields of microelectronics ever since early 1980s. According to a recent global digital signal processor market report in 2012 [93], market decline had been witnessed following the global economic slowdown of 2008-2010. Even so, the market is now recovering, especially for the use of digital signal processors in the automobile industry for manufacturing vehicle parts such as digital radios, voltage regulators, etc. as well as in consumer electronics namely digital cameras, printers, etc.

From the technical perspective, Frantz from Texas Instruments [94] stated that the continued growth of DSP systems will depend on four aspects of technology: the underlying manufacturing processes, the DSP core and chip architectures, the software for development and applications, and most importantly, innovation. On the other hand, DSP techniques are also of great attention, particularly in the realm of sensor networks where many types of wired and wireless sensors are presented as distributed arrays. In such networks, the amount of data recorded for extended period of time would be too much for any person to analyse. Automatic processing is therefore needed. The importance of DSP techniques can then be greatly reflected on how to extract useful and relevant information effectively and efficiently by innovative DSP algorithms, and further analyse and produce sensible results. One recent typical example can be made in the field of

body sensor networks. In this example, a general signal processing and classification framework for wireless medical embedded systems has been proposed [95]. Critical aspects of these systems have already been mentioned, such as preprocessing including data sampling, filtering, signal transformation including segmentation, useful feature extraction, and classification including centralised and distributed data processing. One thing worth mentioning is that, wearable devices favour miniature size and infrequent recharging. Therefore, they are normally constrained by processing capabilities and battery life. Because of these issues, complex signal processing routines are not favoured. Also, because of the considerable power drain in the front-end radio system, it is rather preferable to run simple but effective signal processing tasks on the sensor to avoid unnecessary data transmission. Doing so would greatly reduce the power consumption of communications.

Lastly, advantages and disadvantages of DSP are discussed here. As mentioned before, tasks where analogue methods were used previously or tasks that were difficult or even impossible for analogue methods can now be easily implemented in most cases in the digital domain. First of all, DSP enjoys the flexibility in modifying specific functions of the algorithms at ease. The analogue side, however, has to go through the design flow mostly from scratch down to testing and verification if specifications changes. This is generally the case of *Application Specific Integrated Circuit* (ASIC). Though the Field Programmable Analogue Array (FPAA) is programmable, it is still under development and thus not widely used [96]. Secondly, precision for DSP mainly depends on ADC quantisation resolution, word length, fixed- or floating-point arithmetic, etc. But for the analogue side, much more difficult-to-predict elements are usually involved during design, such as parasitic capacitance and inductance, ambient temperature, technology process variation, etc. All of these make it very hard to control the precision. Thirdly, DSP nowadays can be fairly cheap to design and implement in terms of hardware and manpower. However, analogue could be of vast investment because ASIC may consume significant amount of time, cost and manpower. Fourthly, complex and sophisticated signal processing algorithms are possible for DSP. Digital processors are capable of running functions that are impossible for its opponent, such as linear phase response and adaptive filtering algorithms. On the down side however, speed and bandwidth can be practical limitations for DSP. This is generally because ADC and digital processors are not competent enough for applications that require extremely fast speed and wide bandwidth, such as optical signal processing. Analogue in this case comes with natural superiority of speed and bandwidth.

2.2.2 Preliminaries of Wavelet Transform

2.2.2.1 Introduction

Wavelet Transform (WT) has been a very successful signal processing technique in engineering over the past three decades. Along with rapid advancement, it has been very popular in various kinds of practical applications, including biomedicine [97], communications [98], chemistry [99], data mining [100] etc. Its adaptive, multiresolution capability has rendered itself a powerful mathematical tool for these applications.

However, attracting sufficient attention does take time. Even until the late 1980s, only growth in the theory and practice of wavelet transform had been reported. Since then, both systematic study as well as applications to engineering have developed rapidly. Beginning with Fourier Transform (FT) in 1807, the approximation of a complex function can be calculated as a weighted sum of basis function – sinusoid. However, sinusoids have perfect compact support in the frequency domain, but not in the time domain. Stretching out to infinity in time makes it very difficult to approximate non-stationary signals, e.g. biomedical signals whose frequency content varies along the time. Therefore, to analyse this type of signal is deemed inappropriate.

To address this problem, mathematicians and engineers started to modify FT to support the analysis of non-stationary signals. In 1946, the *Short-Time Fourier Transform* (STFT) was introduced by Gabor [101]. It is a technique that segments the signal into intervals and applies FT on each of them. From the segmentation it is able to provide a true *Time-Frequency Representation* (TFR) of the signals. Since then, many other TFRs have been developed between 1940s and 1970s. Nonetheless, no one had ever looked at an analysis in which high frequency with short-time spans and low frequency with long-time spans were both present in the signal at the same time. It was not until Morlet made his first attempt in late 1970s. From there, different window functions for different frequency bands were introduced by stretching or shrinking the basis function – Gaussian. This was when *wavelet* was first mentioned. Formalisation of transformation and inverse transformation was then achieved by Morlet and Grossman in 1980. To carry the work further, Meyer constructed orthogonal wavelet basis functions in 1984. In 1989, Mallat proposed *Multiresolution Analysis* (MRA) for *Discrete Wavelet Transform* (DWT), where the decomposition of a discrete signal into its dyadic frequency bands was given by a series of high-pass and low-pass filters to compute the DWT coefficients at different decomposition scales [102]. Almost the same time, Daubechies developed the wavelet frames for discretisation and scale parameters of wavelet transform [103]. Research in late 1980s had clearly brought wavelet transform into more noticeable recognition while, more importantly, laying down the foundation of modern wavelet theories and its applications. It also developed the transition from continuous to discrete signal analysis [104].

2.2.2.2 WT Versus FT and STFT

Having described the history of the wavelet, comparison with conventional DSP techniques should be made to introduce the advantages of the wavelet over them[10]. The most popular and conventional one must be FT (or FFT). As is known, it is a technique that transforms the time-domain signal into a frequency-domain signal. The advanced version of this technique is STFT, as we have already discussed. Now, let us consider a signal,

$$x(t) = cos(2\pi 10t) + cos(2\pi 20t) + cos(2\pi 30t) + cos(2\pi 40t) \tag{2.6}$$

This equation is an example of a stationary signal that has 10 Hz, 20 Hz, 30 Hz and 40 Hz, and *stretches* throughout an indefinite time span. FT works perfectly on this type of signal. However, on the contrary, let us consider another signal:

$$x(t) = \begin{cases} cos(2\pi 10t), & t = \quad 1 \text{ to } 250\ ms \\ cos(2\pi 20t), & t = 251 \text{ to } 500\ ms \\ cos(2\pi 30t), & t = 501 \text{ to } 750\ ms \\ cos(2\pi 40t), & t = 751 \text{ to } 1000\ ms \end{cases} \tag{2.7}$$

This signal consists of the same frequency components from the last example, but *scatters* at intervals in the time domain. This effectively makes it a non-stationary signal. According to the Heisenberg Uncertainty Principle [104] in time-frequency information, one cannot know exact information of a signal. In other words, one cannot precisely know what spectral components exist at what instances of time. However, what one can know is the time interval in which a certain band of frequency lies. This eventually refers to a resolution problem. So, the narrower the time interval is in which we use to observe the signal, the better the time resolution while the poorer the frequency resolution will be, and vice versa. Accordingly, it can be seen that FT hardly provides a sensible estimation of frequency components, as in nature it supports no time information. To overcome this, the STFT treats non-stationary signals as stationary ones by dividing them into blocks of short, pseudo-stationary segments with windows. The width of the window is small enough to satisfy valid stationarity. At the expense of perfect frequency resolution, doing so exchanges time information for non-stationary signal analysis. However, resolution is still a big problem. Since the width of the window function such as Gaussian in STFT is fixed during the analysis, resolutions of time and frequency are unchanged regardless of low or high frequencies (as shown in Figure 2.11). This makes it very difficult to observe in the time-frequency domain, particularly when low and high frequency components both exist within the same time interval. Though the window can

[10] More detailed comparison can be referred to [105].

be modified to lend STFT sufficiently to one side of resolution, it will correspondingly lose the other at its expense.

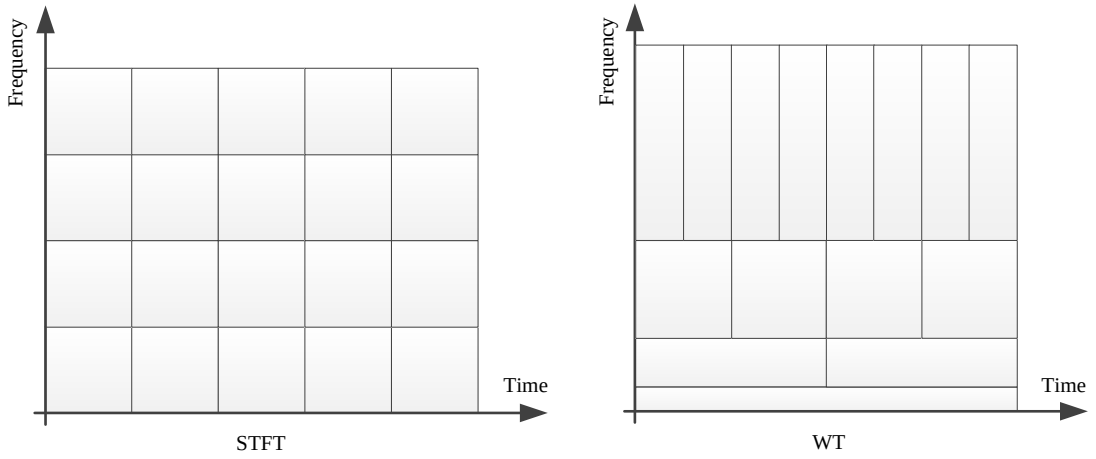


Figure 2.11: Resolution of STFT and WT in time and frequency.

Here comes the WT into the picture. The great advantage of WT lies in handling both time and frequency resolution equally well by the deployment of MRA. Essentially MRA analyses the signal at different frequencies with different resolutions, and every spectral component is resolved unequally in contrast to STFT. This makes WT capable of preserving time information and frequency information at the same time. Such properties render WT a useful tool to present good frequency resolution and poor time resolution at low frequencies, and poor frequency resolution and good time resolution at high frequencies. These can be demonstrated in Figure 2.11, where the shorter height but longer width of the resolution box can be seen at lower frequencies, and the longer height but shorter width of the resolution box can be seen at higher frequencies. This enables effective signal analysis, especially in cases where low frequency components exist for long durations while high frequency ones exist for short durations. Biomedical signals like the ECG fit in exactly the realm of WT applications.

2.2.2.3 Basics of Wavelet Transform

Wavelet transform analysis utilises wavelets, i.e. little wavelike functions, to transform the signal of interest into another representation that is easier to observe and more useful when processing. It is particularly useful for non-stationary signals. Mathematically speaking, WT is a convolution of the wavelet function with the signal. During the convolution, the wavelet can be manipulated in two ways: it can either be moved to various temporal locations of the signal, stretched or squeezed. These two closely relate to the TFR of wavelet transform – a better match of the wavelet function to the shape of the signal results in higher transform coefficients (or larger magnitude) for a specific range of temporal location and frequency in the time-frequency domain, or vice versa. Moreover,

the transform can be done in a smooth continuous fashion with *Continuous Wavelet Transform* (CWT), or in discrete steps with *Discrete Wavelet Transform* (DWT) [106].

In essence, wavelets are mathematical functions. What makes it more useful is that these analysing functions work in accordance with an important idea called *scale*. As is known, in general the analysing functions are mathematical equations that satisfy certain mathematical criteria (e.g. finite energy and admissibility condition[11]). They are used in representing data or other functions by means of approximation using the superposition of functions. The same idea has actually existed since FT was introduced, where sines and cosines can be superposed to represent other functions. Different to FT however, wavelet analysis handles data at different time and frequency resolutions. This effectively renders it advanced to FT thanks to the deployment of scale. In order to understand the concept of scale, let us image a *window* of scale with different sizes – If we looked at a signal through a large window, gross features would be noticed (a higher scale, a larger window and therefore a lower frequency). On the other hand, if we looked at the signal through a small window, smaller features would be noticed instead (a lower scale, a smaller window and therefore a higher frequency). So, it is phrased as *to see both the forest and the trees*, so to speak [107]. This makes wavelets interesting and useful.

The wavelet analysis procedure adopts a wavelet prototype function, called an *analysing wavelet function* or *mother wavelet*. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. To mathematically show how it works, the CWT of a continuous signal $x(t)$ can be shown as follows [108, 104]

$$W(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi^*(\frac{t-b}{a})dt \tag{2.8}$$

where * denotes the complex conjugate notation, and $a > 0$ and b are scale and translation parameters for temporal localisation respectively. $\psi(t)$ is the mother wavelet and $W(a,b)$ is CWT outcome of $x(t)$. From Equation 2.8, it can be seen that WT performs a decomposition of signal $x(t)$ into a weighted set of scaled wavelet functions [109]. It can also be seen that time varying spectral analysis is performed, where scale a effectively plays the role of a local frequency: along with the increase of a, wavelets are stretched and low frequencies are analysed; with the decrease of a, wavelets are contracted and high frequencies are analysed. In fact, this shows how scale is reflected in mathematical terms. Besides, its inverse CWT (ICWT) can also be given as [104]

$$x(t) = \frac{C_\psi}{a^2} \int_{a>0} \int_b W(a,b)\psi(\frac{t-b}{a})da \cdot db \tag{2.9}$$

[11] It requires the wavelet function to be of finite support and oscillatory.

where C_ψ is a constant subjected to $\psi(t)$.

In addition, as an associate to wavelet function, *scaling function* $\phi(t)$ also needs to be mentioned [102]. It is basically a function that covers the low-pass spectrum of wavelet analysis instead of the deployment of an infinite number of wavelets to do the job. By combining these two functions, the spectrum can be handled by wavelet function up to scale j, while the rest can be handled by the scaling function. As we will see, the scaling function in fact works as a low-pass filter and the wavelet function works as a high-pass filter in DWT.

So, to obtain the DWT, parameters a and b must be discretised. By discretising these two as $a = 2^s$ and $b = k2^s$, orthonormal basis functions can be produced. Thus, we have our wavelet basis functions as [107, 108]

$$\psi_{s,k}(t) = 2^{\frac{-s}{2}}\psi(2^{-s}t - k) \tag{2.10}$$

Here, parameters s and k are integers of the power of two that scale (or dilate) the mother function $\psi(t)$ to generate a family of discrete wavelets, for example Daubechies wavelet family. Scale index s indicates the wavelets width, and location index k gives its position.

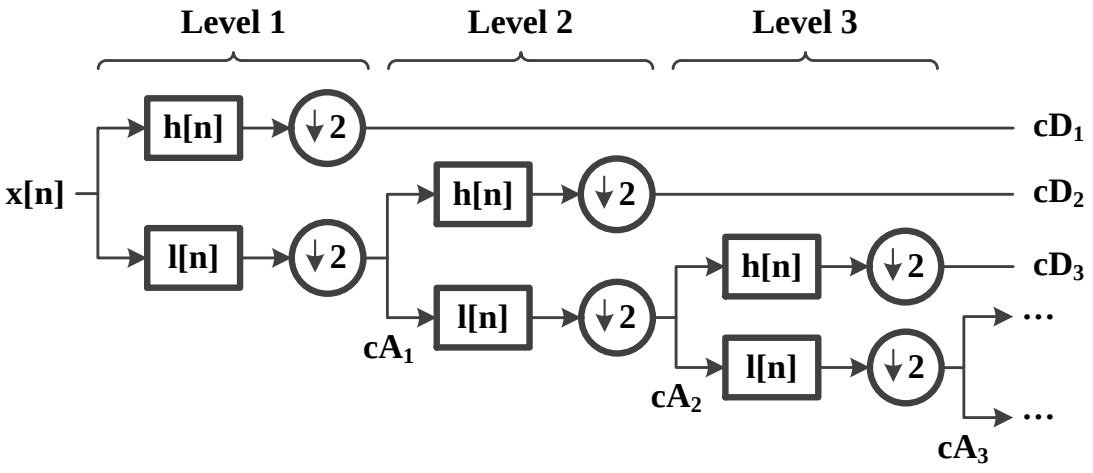


Figure 2.12: Functional block diagram of discrete wavelet transform.

To implement the idea of DWT, [102] shows that DWT of a discrete signal can be obtained by deployment of MRA. They are implemented as high-pass and low-pass filters followed by respective down sampling by two, in an iterative fashion. In other words, cascaded discrete filters can be deployed to perform recursive computation of DWT coefficients, which is contrary to evaluating wavelet coefficients via the integral in CWT. To illustrate the concept, Figure 2.12 depicts the corresponding diagram of DWT.

From there we can see $l[n]$ and $h[n]$ stand for low-pass (scaling function) and high-pass (wavelet function) filters, respectively. At each decomposition level j, detail DWT coefficient cD and approximation cA are produced after filtering and down sampling, which can be given as

$$\begin{aligned} cA_j[n] &= \sum_{k=-\infty}^{\infty} cA_{j-1}[k] \cdot l[2n-k] \\ cD_j[n] &= \sum_{k=-\infty}^{\infty} cA_{j-1}[k] \cdot h[2n-k] \end{aligned} \tag{2.11}$$

2.2.3 Digital Signal Processing for ECG

DSP has a wide variety of applications, including digital filtering, spectral analysis, compression and image processing in various kinds of biomedical signals, like the ECG, Electroencephalogram (EEG), Electromyogram (EMG), Magnetic Resonance Imaging (MRI) image, etc. However, as the main object in our study, we focus more on the ECG as a one-dimensional signal. Having explained DSP in Section 2.2.1, this section will cover generic processing that is tailored to the ECG. The main flow of the discussion below will primarily follow the block diagrams in Figure 2.13[12]. The ECG preprocessing and the ECG signal alteration will be covered briefly with separate focuses on sampling and filtering as well as segmentation and feature detection, respectively.

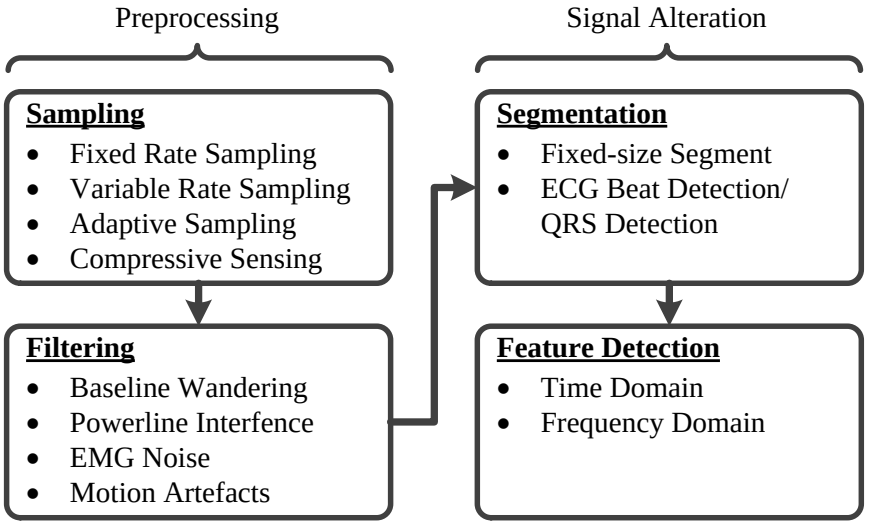


Figure 2.13: Block diagram of general subjects involved in digital signal processing for the ECG.

[12]This figure is inspired by Figure 1 in [95].

2.2.3.1 ECG Preprocessing

At the first stage of any processing, preprocessing is clearly required in ECG applications. That includes sampling that affects the final data volume for storage, data reconstruction, etc., and filtering that cleans up the signals for further processing, including signal transformation, feature extraction and classification. In biomedical signal processing, the signal morphology and the temporal localisation of an event is extremely important, meaning that preprocessing operations are not expected to damage any such relevant information in any ways. Otherwise, the outcome signal would not provide clinically sensible information, often leading to faulty decisions.

Sampling

- *Fixed Rate Sampling*: It is the most common and simple type of sampling. The sampling frequency has to follow the Nyquist sampling theorem. In practice, the frequency is often set to be higher than usually required.
- *Variable Rate Sampling*: Different to fixed rate sampling, this sampling procedure varies with variable sampling rates. Usually it can be manifested as variable sample rate generator sitting next to the ADC [110] and generating different clock signals. These signals are then multiplexed in accordance with the user. Thus it allows the control of the sample rate of the ADC and effectively realises variable rate sampling.
- *Adaptive Sampling*: Adaptive sampling is a technique that automatically changes the sampling rate based on the data. Because the frequency of the biomedical signals generally vary with time, it is possible to reduce the data volume with this technique. An adaptive sampling system for Body Sensor Network (BSN) is proposed in [111]. Also, a signal may be reconstructed from fewer samples than required in traditional sampling methods, just by taking advantage of the spatial properties of the signals with this technique [112].
- *Compressive Sensing*: Even more recent and advanced than adaptive sampling, compressive sensing is a method that can recover certain signals and images from far fewer samples by analysing sparsity and incoherence in the signals [113]. A practical example for biomedical signals in mobile healthcare is given in [114], where packet loss mitigation based on compressive sensing was proposed. It was reported that fewer sensor measurements than suggested by the sampling theorem can be used to recover signals with arbitrarily fine resolution.

Filtering

- *Baseline Wandering*: In ECG signals, especially in ambulatory situations, the ECG baseline tends to wander or fluctuate. A wandering baseline indicates very

low but unwanted frequency contents throughout the signal, usually lying below 0.5 Hz. Low-frequency components of the heartbeat, such as P, T and ST segments, could be affected by this type of noise. Thus removal of the wandering baseline is necessary in this thesis. However, low-frequency variations can also be used in other applications, e.g. Cardiopulmonary Coupling (CPC) detection, where ECG-derived respiration is needed to measure the coupling between respiration and heart rate [115, 116].

To handle baseline wandering, linear filtering and polynomial fitting are commonly used. For linear filtering, time-invariant and time-variant filtering techniques can be applied. In the former, cut-off frequency and phase response characteristics of the filter are crucial. In this case, linear phase high-pass filtering with 0.5 Hz cut-off frequency is generally preferred in order to prevent phase distortion from altering ECG wave components' properties while removing the wandering baseline [117]. In the later, a time-variant approach is capable of coupling the cut-off frequency to the prevailing heart rate instead of a fixed one. It provides better baseline removal, especially at low heart rates or excessive baseline wandering [118]. In addition to linear filtering, polynomial fitting is also commonly used where estimations of the baseline are done by passing through representative samples (also called *knots*) of the ECG. Thereby baseline wandering can be removed by subtracting the estimated baseline from the original ECG. In particular, third-order polynomial (i.e. cubic spline) fitting has shown good performance, especially in tracking rapid wandering baselines as we tend to have more knots [119].

Apart from the above techniques, two other ways of removing wandering baselines are also available: DWT and median filtering. DWT can be applied as a filter bank to the signal, and from there DWT coefficients of low-frequency components can be forced to zeroes. This way we can effectively filter out unwanted wandering baselines. An example can be found in [120]. On the other hand, median filtering is capable of removing P, QRS and T waves, which eventually results in a pure wandering baseline. With it, the original signal can be subtracted and thus removal can be achieved. Bear in mind that, an appropriate window size for the filter has to be chosen carefully, as otherwise signal quality could be seriously jeopardised [121].

- *Powerline Interference*: Electromagnetic fields caused by a powerline represent a common noise source in the ECG. Such noise is characterised by 50 or 60 Hz sinusoidal interference. Having this type of noise would make the ECG analysis difficult, particularly when detecting high-frequency features in the time-domain. As a result, different techniques are commonly used to handle it, for example linear and nonlinear filtering, estimation-subtraction, etc. Firstly, a linear bandstop filter can be used, with a notch at specific frequencies. Increasing the orders of the filter can lead to a narrower notch. However, it may cause problems in

having increased transient response time of the filter and thus spurious waveforms may be introduced. Secondly, in an attempt to be less sensitive to transients in linear filtering, nonlinear filtering can be utilised in this case. A method and an example can be found in [122]. Thirdly, estimation-subtraction can be viewed as an approach where the amplitude and phase of the interfering sinusoid in an isoelectric segment is estimated, followed by subtraction of the estimated sinusoid within the entire heartbeat. [123] reported that such a technique can reduce powerline interference effectively without introducing any additional noise.

- *EMG Noise*: It is interference generated from electrical activity of the muscle during muscle contractions. This type of noise contains frequency contents that spread through the ECG frequency spectrum, making it extremely difficult to handle. Because of that, it may be even impossible to reconstruct the original noise-free signal. To reduce the noise effect, a wavelet threshold method was reported to be efficient [124]. Time-varying lowpass filtering using a filter with a variable frequency response was also suggested [125].

- *Motion Artefacts*: Motion artefact is the noise that results from motion of the electrode in relation to the patient's skin. Similarly to EMG noise, its frequency contents overlap those of the ECG. The ECG may in some cases be corrupted seriously enough to be clinically unusable. Therefore, methods have been developed to tackle this problem, such as adaptive filtering with a reference signal to either motion noise correlated in some way with the noise in the noisy ECG, or a signal correlated only with the noisy ECG [126, 127].

2.2.3.2 ECG Signal Alteration

Signal alteration includes segmentation and feature detection. In a way, specific events (or discrete time intervals of the signal) are demanded, upon which information processing algorithms such as feature detection and classification will be executed. This is when segmentation is needed. Segmentation algorithms divide the data stream into separate and clinically meaningful segments of interest so that a reasonable ECG heartbeat can be obtained. On the other hand, feature detection algorithms are dedicated to detecting/extracting features of clinical importance in an automatic fashion. Apart from a direct approach based on the time domain, and popular signal processing techniques like FT, WT are often used to transform the signals into frequency or time-frequency domains in order to process and find more valuable features. Features obtained are commonly applied in applications such as classification, visualisation, compression, and so on.

Segmentation

- *Fixed-Size Segment*: it is a computationally simple method that divides the long-duration ECG series into fixed-size segments. Since consecutive ECG heartbeats may vary in terms of intervals, a more advanced approach might be to apply a template of a reference ECG heartbeat to map and match possible heartbeats along the series.

- *ECG Beat Detection/QRS Detection*: ECG beat detection, or primarily the detection of the R wave within QRS complex, generally consists of four blocks: linear filtering, nonlinear filtering, peak-detection logic, and decision. Since typical frequency components of a QRS complex range from 10 Hz to about 25 Hz, filtering is commonly deployed to attenuate other signal components that are not necessarily needed during detection, for instance the P and T wave, a wandering baseline and noise. This is followed by peak detection to capture the possible R peak and to rule out the false-positive by the decision block [128]. Perhaps the most well-known example is Pan & Tompkins algorithm. It is an algorithm that utilises the slope, amplitude and width of the ECG signals [129]. The advantage of the algorithm is that the filters involved are not computationally expensive and the detection logic is quite simple, making it fairly suitable for real-time implementation [95]. More detailed principles of other software algorithms, such as wavelet-based detection, neural network, and adaptive filters are reported in [128]. An evaluation of detection accuracy as well as a rough comparison of the computational complexity between them are also presented. Apart from that, a very detailed review on QRS detection methodologies for wearable battery-powered wireless ECG systems is given in [130]. From there, a comparison of QRS enhancement as well as detection techniques based on associated algorithm principles and assessment criteria (robustness to noise, parameter choice, computational complexity) are stated clearly. The conclusion of this work has suggested adaptive thresholding for QRS detection, though such a setting of multiple empirical thresholds might lead to certain inaccuracies, especially in arrhythmias and noisy QRS complexes.

Feature Detection

- *Time Domain*: As time series that explicitly manifest the cardiac events of the heart, the ECG exhibits various morphologies that correspond to different heart conditions. That also means that a lot of clinically useful time domain features can be observed in the ECG. These features include but are not limited to the morphology of ECG waves, the duration of waves, the RR interval, the area under the curve, and the angles of lead vectors in the VCG, etc. To process and use them to our advantage, automated algorithms are commonly applied. Extensive research has been done on feature detection, a feature selection and classification, all of which involve this key processing as a preceding step. For instance, one well-known work was published based on the Pan & Tomkins work [129] for

measuring clinical features for multilead ECGs [131]. Another similar work with a more sophisticated scheme combining extrema detection and slope information with adaptive thresholding was conducted in [117]. In addition, typical automated classifications of heartbeats using morphology and interval features of the ECG were also reported in [121, 132]. In these works, extensive use of time domain features was made to cover a wide range of features, coupled with a linear discriminant model to classify ECG heartbeats. Quite recently, features based on the spatial analysis of the VCG was obtained, following which a supervised learning model was constructed to classify the presence of scar tissue in the myocardium [133]. Besides, there was one international cooperative project called Common Standards for Quantitative Electrocardiography (CSE) [134] initiated in early the 1980s, with the aim of standardising computer-derived ECG measurements. Since then it has been regarded as a reference for future works on ECG feature detection.

- *Frequency Domain*: An alternative to time-domain analysis is frequency domain. However, unlike the time domain, the frequency domain is capable of unveiling the frequency and phase spectrum of the signal. This allows a different perspective of analysing temporal morphologies, making it possible to find features that are not obvious in the time domain. Research in this area can be given as follows. First, significant number of FT-based spectral features were investigated in [135, 136], where selection of optimal parameters was made to perform ECG diagnostic classification. This effectively demonstrated that researchers were able to extract features in the frequency domain for further applications. Later when WT began to attract researchers' attention, this time-frequency domain technique was used in many ECG applications. The most well-known ones on the delineation of ECG fiducial points are [137, 138], where a quadratic spline wavelet was used and demonstrated excellent performance in a standard MIT/BIH database. In addition to delineation, wavelet transform was also used to produce coefficient sequences at different scales, followed by optimisation or further processing (e.g. auto/cross-correlation between leads), and acted as inputs to classification algorithms to execute prediction on various types of heartbeats [139, 140, 141].

2.3 Machine Learning

2.3.1 Introduction

Machine learning (ML) is one of the major branches of Artificial Intelligence (AI) and indeed considered to be one of the most rapidly developing subfields of AI research. It is a framework of processing data and using the already-existing data (either labelled or not labelled with classes) to learn/infer a functional relationship between a set of

attributes (or features) and associated responses (or classes), in order to predict the unseen data [142]. On a more abstract level, it is a concept of programming computers so that a performance criterion using example data or past experience may be optimised. Learning is needed when a computer program cannot be directly written to solve a given problem. Learning is especially necessary when human expertise does not exist, or when humans are not able to explain their expertise; it is also necessary when the problem changes through time, or relies on the particular environment [143]. Bear in mind that, whatever conclusion drawn from our ML analysis in fact depends on the dataset provided. According to the *No Free Lunch Theorem*, the best classification algorithm is not at all available [143]. Given a dataset, certain algorithms exhibit very promising performance; when replaced with another dataset, they could bring a disastrous performance. A learning algorithm is defined as *good* only in the sense that its intrinsic mechanics suit well the properties of the data.

Historically speaking, the development of ML bagan in 1957 [144]. It was at the time that the perceptron model was invented as an initiative in *neural networks* [145]. Over the next decade, it became very popular until the limitations of this model in expressing complex functions was pointed out in 1970. ML had been dormant for sometime afterwards, and expert systems had become the mainstream approach in AI. Fortunately, the *decision tree* model was invented in mid-1980s [146]. It is a technique that is simple, easy-to-understand, and can be viewed by a human very easily. At nearly the same time *multilayer neural networks* were also invented [147], which is a technique that is able to express any function with sufficient hidden layers. Both revived ML as a popular research topic again. Because of the invention of the World-Wide-Web as well as the demand of big-data based applications, people call for more sophisticated methods in automation for data analysis. Since then ML has been thriving rapidly. In 1995, the *Support Vector Machine* (SVM) was invented and soon enough became one of the most popular techniques in ML community [148]. After 2000, *logistic regression* was re-designed for large-scale ML problems and found to be a practical algorithm in many large-scale commercial systems such as text classification [149, 115, 150]. In addition to the methods described above, the ensemble based system, mainly started in 1990, has also attracted considerable attention due to its better predictive performance when combining multiple models over just one, and this technique has been growing since then [151]. Deep learning, a term that attracted public attention again after [152] published in mid-2000s, has become part of the many state-of-the-art techniques in machine learning, particularly in applications such as computer vision and automatic speech recognition.

Having briefly considered the history of ML, let us focus on its subdivisions. Technically speaking, ML can be divided into several fields. Here, we mainly cover those relevant to our work[13]. Depending on the data and associated labels available, learning can

[13] Apart from those relevant to our work, Bayesian methods are also used extensively in the field of machine learning in general [153] as well as specific areas like ECG processing [154, 155].

be categorised into *supervised learning* and *unsupervised learning*[14]. Apart from the learning part, *dimensionality reduction* is also covered here. Note that in this thesis, classification in supervised learning and feature selection in dimensionality reduction will be primarily discussed and deployed.

- *Supervised Learning*: When the training data are labelled with classes, they can be used to train learning algorithms, thereby predicting new ones. A decision boundary is thus trained based on the available data, and hence is used to separate out the different classes. A wide range of techniques were introduced to support supervised learning, for instance discriminant analysis, SVM, k-Nearest Neighbour (k-NN), decision tree, Naïve Bayes, etc.
- *Unsupervised Learning*: Given data that are not labelled with any classes, unsupervised learning can be applied to find clusters of similar data, and predict new ones using these clusters. Approaches to this type of learning include k-means clustering, mixture models, etc.
- *Dimensionality Reduction*: As its name implies, dimensionality reduction reduces the current dimension of the feature space into a lower one. Three different ways can be applied [156]: (1) feature selection, in which useful features are intelligently chosen out of the available ones; (2) feature derivation, which apply transforms to derive new features from the old ones; and (3) clustering, which groups together similar datapoints and see which features may be used. Lots of algorithms are available in dimensionality reduction, particularly in feature selection. The reader can refer to [157] for more details.

With the different kinds of methods listed above, ML has been enormously useful in many applications. These include for example, commercially available systems for speech and handwriting recognition; learning customers' behaviour to improve the quality of management in retail companies; large amount of data mining and analysis in bioinformatics, so on and so forth. In terms of tools, a lot of open-source software for machine learning and pattern recognition is available. Perhaps the most famous ones are WEKA (a collection of machine learning algorithms for data mining tasks) [158], Pattern Recognition Tools (A Matlab toolbox for pattern recognition) [159], libSVM (A Library for Support Vector Machines) [160]. Commercial software is also available, for instance Statistics Toolbox in Matlab, SPSS, etc. Also note that, *data mining* as a commonly used term nowadays should not be confused with machine learning. Data mining is defined more towards discovering interesting patterns in large data sets, while ML is defined more towards the actual learning methods applied on the data [161]. Although data mining may be used throughout the text (mainly out of the respect for the literatures we reference), the difference between the two should be noted.

[14] In fact, there are two more learning types: reinforcement learning and evolutionary learning. But they are not covered here. The reader can refer to [156] for more information.

2.3.2 Background in Biomedical Engineering

A large number of applications benefit from the development of ML, such as general signal processing, telecommunications, robotics and dynamic control, financial or other time series data analysis, etc. Among them, biomedical engineering has been one of the most active research and application subjects. Major machine learning applications in biomedicine can be divided into two streams: medical diagnostic reasoning and biomedical signal processing [162]. The former application involves expert systems and ML model based schemes, providing mechanisms for the generation of hypotheses from patient data. The latter involves modelling the nonlinear relationships between data to help find essential features and information hidden in the physiological signals, whose characteristics are not easily manifested. In fact, ML offers powerful methods and tools to facilitate solutions in diagnostic and prognostic problems in various biomedical domains. For example, (1) computational neuroscience is one of the rapidly expanding areas, where new models of neurons, local neural circuits as well as new learning rules are proposed to help understand how the nervous system actually works; (2) bioinformatics and genomics is another area, where ML algorithms are evaluated in gene sequence analysis, development of artificial immune systems, genomic data mining, as well as biometric identification, image processing and handwritten character recognition, etc; and (3) the analysis of biological signals for detection or identification of certain pathological conditions. Sophisticated methods are used to extract features from the ECG so as to diagnose various cardiovascular disorders, including arrhythmias and heart rate viability, as well as features from EEG signals in order to detect or diagnose various neurological conditions [151].

Nowadays, since there are relatively inexpensive ways to collect and store medical data (which are shared in large information systems), machine learning technology is currently well-suited for analysing medical data. That leads to a fact that, patient records with availably known correct diagnoses can simply act as input to run a learning algorithm. After that, medical diagnostic knowledge may be derived from the description of cases, which has been solved in the past, in an automatic fashion. Once it is done, the classifier derived may be deployed in two directions: to train students, physicians or non-specialists to diagnose patients in a special diagnostic problems, or to support the physician when diagnosing new patients so as to improve the diagnostic speed, accuracy and reliability [163]. The latter is particularly useful in scenarios where patients are remotely monitored and care can be delivered once an alarm is raised.

Remote monitoring in particular, has been attracting the attention of ML researchers for the past few years. A very recent review paper on data mining in health monitoring systems, with a specific concern in wearable sensors, comprehensively states the latest methods and algorithms deployed when performing data analysis of vital signs captured from wearable sensors in healthcare services [164]. Specific focus has been put

on the common data mining tasks applied on anomaly detection, prediction and decision making, as well as the suitability of these methods when processing physiological data. Moreover, it also provides a list of guidelines for the selection of data mining methods, as well as the general challenges that are faced in data mining in health monitoring.

2.3.3 Feature Extraction

In ML, feature is very important in that it represents the informative properties of the subjects we want to learn about. Therefore, the extraction of features is deemed to be the key. *Feature extraction* is a process of finding the most informative and yet compact set of features, so that the efficiency of ML tasks, data storage and data processing can be improved. More importantly, the most common and convenient ways of representing data for any classification and regression problems are feature vectors, which are originally defined in feature extraction. So, this makes feature extraction a unique research topic, and its knowledge is commonly shared by ML, data mining and also fuzziness and soft computing [165], etc.

Feature extraction consists of two main components: *feature construction* and *feature selection*. As their names suggest, feature construction refers to constructing informative representations of the data, while feature selection, on the other hand, stands for selecting a subset of relevant features. Furthermore, the benefits of doing feature extraction lie in two facets: feature construction builds up and calibrates the data for subsequent statistics or ML algorithms, and feature selection effectively reduces training and utilisation times, defies the curse of dimensionality [166] to improve prediction performance, reduces the measurement and storage requirements, and also facilitates data visualisation and data understanding [167, 168]. Note that, the term feature extraction in other fields like signal processing should not be confused with the term in ML. That is because in ML feature extraction refers to a broader range of alteration comprising feature construction and feature selection, whereas other fields normally refers to it as detection and delineation of local features.

2.3.3.1 Feature Construction

Raw data can be of many types. It can be nominal, binary, ordinal, numeric, discrete or continuous [161]. Some of the data can be used for prediction analysis straightaway, while some of them cannot. Therefore, converting raw data into a set of useful features usually requires some preprocessings (i.e. feature construction) before we feed the features into a feature selection or even a prediction module. To describe the preprocessing steps, some notations should be introduced. Let $\mathbf{x}$ be a feature vector of dimension n, $\mathbf{x} = [x_1, x_2, \cdots x_i]^T$. x_i is a component in dimension i. So, the preprocessing transformations may include [168, 169]:

- *Standardisation*: When features originally come from different scales or even units, action can be made to coordinate them onto the same scale. Assuming that the data complies with Guassian law, standardisation can be used upon the data following

$$x_i^{'} = \frac{x_i - \mu_i}{\sigma_i} \tag{2.12}$$

  where μ_i and σ_i are the mean and standard deviation of feature vector $\mathbf{x}$, respectively. Doing so makes all features complied with zero mean and unit variance.

- *Normalisation*: Unlike standardisation, normalisation scales all numeric variables in the range of $[0, 1]$ or $[-1, 1]$, following

$$x_i^{'} = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} \tag{2.13}$$

  where $x_{i,max}$ and $x_{i,min}$ represent the maximum and minimum of feature vector $\mathbf{x}$, respectively.

- *Signal Enhancement*: It generally refers to enhancing the signal-to-noise ratio of the signal or image by applying proper filtering techniques. Operations include de-noising, baseline or background noise removal, smoothing, etc.

- *Extraction of Local Features*: Depending on the specific domain, extraction of local features commonly exploits automatic feature extraction algorithms to detect and capture useful features. These features can be of many types: sequential, spatial or structured.

- *Linear and Non-Linear Space Embedding Methods*: To reduce the dimensionality of the data, especially when it is high, techniques may be used to project or embed the data into a lower dimensional space without losing much information. Here classical methods include Principal Component Analysis (PCA). After reduction, the data can be presented as a lower dimensional feature space for prediction, or simply for data visualisation.

Throughout the thesis, several points from above will be applied where necessary. Notably, the extraction of local features primarily refers to the feature detection of ECG fiducial points, which will be discussed in Chapter 3. Features extracted there will go through feature selection before running classification. Note however that issues may rise when deciding which one should be chosen between standardisation and normalisation when data calibration is needed. Both methods have their drawbacks. Standardisation may not be applicable in some cases when the distribution of the data does not simply follow Guassian; whereas normalisation may not be suitable when regular data are scaled improperly into a very small portion when outliers exist, which is quite common in practice.

2.3.3.2 Feature Selection

In ML applications, thousands or even millions of features could be generated from a system upon which classification is expected to operate. Are all these features interesting? Typically not. In fact, only a small portion of them would appeal to the data analyst. So here comes the feature selection into the picture. Feature selection is a process of selecting a subset of original features in accordance with certain criteria. It is a technique that is important and frequently used in dimensionality reduction for ML and data mining [170]. Unlike other dimensionality reduction techniques such as PCA, feature selection does not alter the original representation of the features, but selects a subset of them. This is important because it significantly facilitates the reduction of features, removal of redundant, irrelevant or noisy data, and thus eventually speeds up the prediction algorithm as well as improving its performance in predictive accuracy and result comprehensibility [171].

At the core of feature selection, the assumption is that irrelevant features are expected to be removed and relevant ones are expected to be preserved. The question then becomes: what are relevant features? To answer this, a definition of relevance and redundancy should be given. Let $\mathbf{f}$ be the full set of features, and f_i be a feature. Also let $\mathbf{S}_i = \mathbf{f} - f_i$, C denote the class label, and let p denotes the conditional probability of the class label C given a feature set. So, the statistical relevance of a feature can be formalised as [172]:

Definition 2.1. A feature f_i is strongly relevant iff

$$p\big(C|\ f_i, \mathbf{S}_i\ \big) \neq p\big(C|\ \mathbf{S}_i\ \big) \tag{2.14}$$

A feature f_i is weakly relevant iff

$$\exists \mathbf{S}_i^{'} \subseteq \mathbf{S}_i, \text{such that } p\big(C|\ f_i, \mathbf{S}_i^{'}\ \big) \neq p\big(C|\ \mathbf{S}_i^{'}\ \big) \tag{2.15}$$

Otherwise, the feature f_i is irrelevant.

Equation 2.14 defines strong relevance, which implies that the feature is indispensable in the sense that loss of prediction accuracy would definitely occur once it is removed. Similarly, Equation 2.15 defines the weak relevance, suggesting the feature may or may not contribute to the prediction accuracy. Moreover, the definition of relevance essentially implies two reasons for having a statistically relevant feature: either the feature is strongly correlated with the class, or a feature subset can be formed with other features and the subset is strongly correlated with the class [171].

Definition 2.2. A feature f_i is redundant iff

$$p\big(C|\ f_i, \mathbf{S}_i\ \big) = p\big(C|\ \mathbf{S}_i\ \big), \text{but } \exists \mathbf{S}_i^{'} \subseteq \mathbf{S}_i, \text{such that } p\big(C|\ f_i, \mathbf{S}_i^{'}\ \big) \neq p\big(C|\ \mathbf{S}_i^{'}\ \big) \tag{2.16}$$

Equation 2.16 states that feature f_i may be categorised as redundant due to the existence of other relevant feature. If only a portion of the relevant features are available, feature f_i may account for prediction improvement [171]. In addition, the averaged correlation among all possible feature pairs within the feature subset after performing feature selection (i.e. redundancy rate) is expected to be small. Otherwise a large correlation value would mean many features selected in this feature subset are strongly correlated. That means redundancy still exists, indicating that the optimal subset of features has not yet been achieved [171].

To clarify these two definitions, let us assume we have a Boolean feature set $\mathbf{f} = \{f_1, f_2, f_3, f_4, f_5\}$. By nature, f_2 and f_4 are negatively correlated, and so are f_3 and f_5. The target class/output equation[15] of C satisfies $C = f_1 \oplus f_2$. Substituting f_2 with f_4, we have $C = f_1 \oplus \overline{f_4}$. According to Definition 2.1, f_1 is strongly relevant, or indispensable; f_3 and f_5 are irrelevant; and f_2 and f_4 are weakly relevant as one of them can be disposed. As for redundancy, assume that we have two Boolean feature sets $\mathbf{f}_1 = \{f_1, f_2\}$ and $\mathbf{f}_2 = \{f_1, f_2, f_4\}$ under the same setting. According to Definition 2.2, $\mathbf{f}_2$ contains redundancy as either f_2 or f_4 can be categorised as a redundant feature due to the existence of the other, whereas $\mathbf{f}_1$ can not.

In summary, the ultimate goal of feature selection is to improve the prediction accuracy by finding relevant features while reducing the overall redundancy.

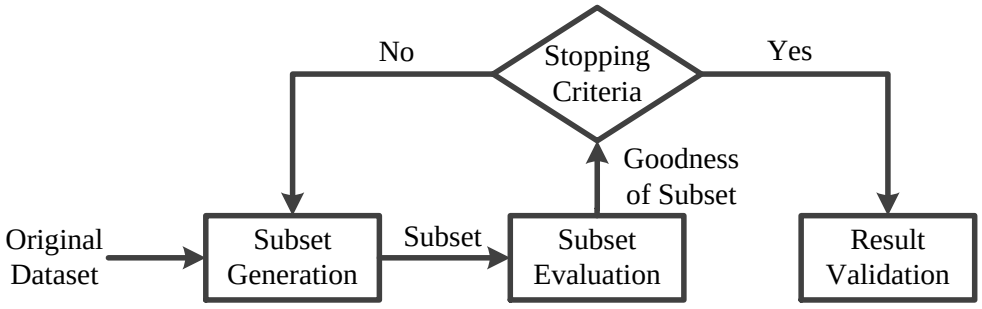


Figure 2.14: Four key blocks of feature selection.

Having surveyed these two concepts, the general procedure of feature selection should now be introduced. Basically four components are present, as illustrated in Figure 2.14[16]: subset generation, subset evaluation, stopping criteria, and result validation [173, 157].

- *Subset Generation*: It is a process of heuristic search in which a candidate feature subset is produced and fed for subset evaluation. Furthermore, two basic elements must be determined in this block: (1) the search starting point, which can be started with an empty set that continuously adds features (i.e. forward search),

[15] $\oplus$ denotes XOR.
[16] This figure is inspired by [157].

or a full set that continuously removes features (i.e. backward search), or even both directions, i.e. adding and removing features in parallel (i.e. bidirectional); and (2) search strategy, which can be categorised into complete search, sequential search and random search. Notice that a further discussion on these strategies will be made in Chapter 5.

- *Subset Evaluation*: In this block, each subset produced from a generation block is evaluated by an evaluation criterion. Depending on whether or not learning algorithms are involved, evaluation criteria can be broadly categorised into two groups: independent criteria and dependent criteria. Within the category of independent criteria, subgroups can be identified which includes (1) distance measures; (2) information measures; (3) dependency measures; and (4) consistency measures. Note that, if a new subset is found to be better under certain criteria, it substitutes the best from the previous one.

- *Stopping Criteria*: With stopping criteria, the loop of finding optimal subsets will be interrupted once the rules are satisfied. Frequently used criteria include (1) no better subset is produced when adding or removing a feature; (2) a given bound is reached, for instance a maximum number of iterations, or a maximum or minimum number of features; (3) a sufficiently good subset is found; and (4) the entire search is finished.

- *Result Validation*: Once the final selected subset is achieved, it is then subject to result validation by some given learning algorithms. Here the commonly used assessment indices are accuracy, redundancy rate, etc.

Now, apart from the general procedure of feature selection, there is one more important concept in feature selection. Depending on how a feature selection search is integrated with learning algorithms, feature selection algorithms can be broadly classified as: *filter* methods, *wrapper* methods, and *embedded* methods [174].

- *Filter*: In this type of method, the relevance of features is evaluated by only taking into account the intrinsic properties of the data itself. Once relevance-related score associated with each feature is produced and ranked, the lowest ones are removed, leaving the relevant ones. The advantages of using the filter have several folds. Among them, high-dimensional datasets can be handled very easily in terms of complexity; operations are computationally simple and fast; and also it is independent of classification algorithm, making it easily applicable to the evaluation of various classifiers once it is done. On the other hand, as it may already show, one major disadvantage of the filter method is the lack of interaction with learning algorithms. Since most filter techniques are univariate and therefore no feature dependencies are considered, prediction performance may be fairly bad in comparison with other types of feature selection methods.

- *Wrapper*: Unlike the filter method, the wrapper method tries to *wrap* around the learning algorithm with a search strategy. In other words, the whole process of finding a good feature subset is completely tailored to one specific learning algorithm. The evaluation of feature subsets is done by training, testing that specific algorithm, and taking its performance as evaluation metric. The advantages of the wrapper method include the interaction between the search of feature subsets and model selection, and also the embrace of feature dependencies. However, the disadvantage is inevitable – it is computationally more expensive than the filter method and risks the problem of overfitting.

- *Embedded*: Attempting to integrate both methods above, the embedded method incorporates the search of optimal subsets into the construction of a learning algorithm. Generally the complexity taken is between filter and wrapper methods. However, only a few algorithms are available in the research community.

Each method listed above has its own strengths and drawbacks. Depending on the applications and context, a suitable one should be carefully considered and then applied. Due to the fact that a large number of algorithms are available in the community, it is quite difficult to make the right choice. To tackle this, in [157] a unifying platform is proposed, with the hope of recommending the most suitable feature selection algorithm among many for a given application. As we will see later, both methods of feature selection and the platform will be successfully integrated into our work: the concept of the wrapper method will be primarily used in Chapter 4, and the filter method as well as the unifying platform will be effectively deployed in Chapter 5.

Finally, feature selection has contributed significantly to various application fields and extensive research has been done, including image retrieval [175, 176], text categorization [177, 178, 179], customer relationship management [180], etc. In particular, a comprehensive review covering feature selection techniques specifically tailored to bioinformatics, available general-purpose softwares as well as dedicated ones, etc., is given in [174]. Besides, a dedicated feature selection repository for research can be found in [171]. It essentially covers a large range of introductory description of the most popular feature selection algorithms and, more importantly, a professional platform of algorithm collections for operating various feature selection algorithms straightaway.

2.3.4 Classification Models

2.3.4.1 General Information

Since the main target of our work is to classify normal and abnormal ECG heartbeats, classification is paramountly important to us in terms of its application and theoretical background. In our database setting, the class information of our patients is directly

available to us. Logically enough, this lends us to classification under supervised learning category.

Now, having discussed supervised classification, learning algorithms within this scope should be carefully considered. First of all, as will be mentioned later in Chapter 4, the number of data samples (i.e. subjects of patient) we consider in our study is quite limited, totalling to around 100. In our thesis, we opt for nonprobabilistic methods. However, we also need to embrace the fact that Bayes methods are also good candidates for classification when having a small number of data [181, 182]. Nonprobabilistic methods are a collection of algorithms that effectively assigns a class to a new object using certain strategies, instead of outputting a probability. Within this category, we have opted for parametric classifiers like discriminant analysis model, SVM, as well as nonparametric classifiers like k-Nearest Neighbours (k-NN). In this thesis, we will be covering different sides of these nonprobabilistic methods, with the hope of demonstrating nonprobabilistic classification and the difference between them, as well as letting them compete for the optimal performance for our purposes.

2.3.4.2 Basics of Our Selected Classifiers

In this section, we discuss the preliminaries of the selected nonprobabilistic classifiers we used throughout our study. In fact, there are different types of classifiers used to solve different classification problems over the last decades [121, 132, 183, 184, 185, 186, 187]. In our particular case, the resource-constrained nature of remote monitoring ECG systems implies that the employed classification methods must be less computationally demanding in terms of the number of arithmetic operations required for labelling, i.e. assigning a class to a test sample. As a result, we have selected to investigate Linear/Quadratic Discriminant Analysis, because they are considered to be computationally efficient [188]. On the other hand, SVM [185] and k-NN [187] have been previously considered for ECG classification in the relevant literature, therefore we have also included them in this study.

Linear/Quadratic Discriminant Analysis There are two ways of explaining discriminant functions for classification. One is *likelihood-based classification*, and the other is *discriminant-based classification* [143]. As we will see later, we opt for the latter as the main theoretical explanation of LDA and QDA, and directly apply the knowledge in our experiments.

For instance, in the case of likelihood-based classification, we first estimate the prior probability $\hat{P}(C_i)$ as well as the class likelihood $\hat{p}(\mathbf{x}|C_i)$, then posterior density is calculated with Bayes' rule. Here C_i refers to the label of class i. Eventually the discriminant function in terms of the posterior probability is derived, which is

$$g_i(\mathbf{x}) = log\hat{P}(C_i|\mathbf{x}) \tag{2.17}$$

Given Equation 2.17, one can assign C_i to a testing sample that exhibits the maximum posterior probability [189]. In other words, one should choose C_i if $g_i(\mathbf{x}) = \max_{j=1\ to\ K} g_i(\mathbf{x})$, where K is the maximum number of class labels.

On the other hand, we also have discriminant-based classification, where a model is assumed directly for discriminant without any estimation of likelihoods or posteriors. In this case, $g(\mathbf{x}|\Phi)$ is the discriminant function with a set of parameters Φ. The goal here is to optimise these parameters to maximise the quality of separation, or the classification accuracy, on a given labelled training set. It is essentially contrary to likelihood-based methods in which sample likelihoods are maximised by searching for optimal parameters. Thus, in discriminant-based classification the concern is not the density estimation of class regions. Rather, it is about the correct estimation of the boundaries between the class regions [143].

As the simplest case of discriminant-based classification, the linear discriminant function for class i can be found as

$$\begin{aligned} g_i(\mathbf{x}|\Phi_i) &= g(\mathbf{x}|\mathbf{w}_i, w_{i,0}) \\ &= \mathbf{w}_i^T\mathbf{x} + w_{i,0} = \sum_{j=1}^{d} w_{i,j}x_j + w_{i,0} \end{aligned} \tag{2.18}$$

Here, $\mathbf{w}$ refers as a weight vector and w_0 as its intercept, together constructing the hyperplane for separation. More importantly, $\mathbf{w}$ determines the orientation of the hyperplane, and w_0 determines the location of the hyperplane with respect to the origin. Next, taking two-class classification as an example, the discriminant function in this case would be

$$\begin{aligned} g(\mathbf{x}) &= g_1(\mathbf{x}) - g_2(\mathbf{x}) \\ &= (\mathbf{w}_1^T\mathbf{x} + w_{1,0}) - (\mathbf{w}_2^T\mathbf{x} + w_{2,0}) \\ &= (\mathbf{w}_1 - \mathbf{w}_2)^T\mathbf{x} + (w_{1,0} - w_{2,0}) = \mathbf{w}^T\mathbf{x} + w_0 \end{aligned} \tag{2.19}$$

and the output label can be given as

$$C_i = \begin{cases} C_1, & \text{if } g(\mathbf{x}) < 0 \\ C_2, & \text{otherwise} \end{cases} \tag{2.20}$$

To help understand Equation 2.19, an illustrative instance of a two-dimensional two-class classification is presented in Figure 2.15[17]. It can be seen that a one-dimensional hyperplane mathematically manifests itself as $g(x) = w_1x_1 + w_2x_2 + w_0 = 0$. Furthermore, it effectively divides the entire feature space into two parts: $g(\mathbf{x}) < 0$ for the negative part and $g(\mathbf{x}) > 0$ for the positive part. That means, given this particular hyperplane (or decision boundary), any new testing sample falling into negative part will be classified as C_1, or C_2 otherwise. In addition, linear discriminant classification also lends itself well to multiple-class situation, under the same assumption that all classes are linearly separable.

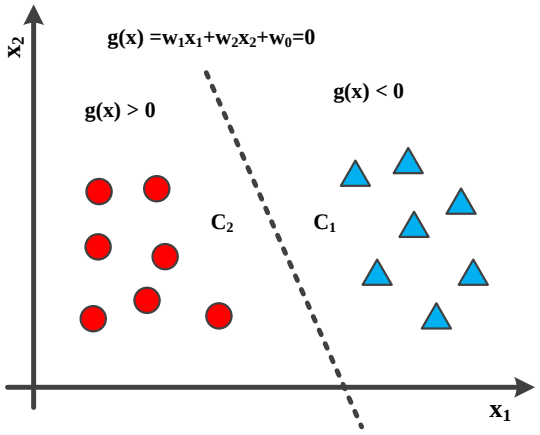


Figure 2.15: An illustrative example of two-dimensional two-class linear discriminant classification.

From a learning perspective, we also need to understand how the discriminant function, or more specifically the weight vector $\mathbf{w}$ and the intercept w_0, is determined. In fact, in likelihood-based classification, parameters are optimised and determined primarily by prior probability $\hat{P}(C_i)$ and class likelihood $\hat{p}(\mathbf{x}|C_i)$, and the fundamental method behind is maximum likelihood. Interestingly, in our discriminant-based approach, the optimisation scheme of these parameters is different – the goal is to minimise the classification error on the training set instead. That means, when $E(\mathbf{w}, w_0|\mathcal{D})$ denotes the error function (or cost function) and $\mathcal{D} = \{(\mathbf{x}_n, C_n)|n = 1, \cdots, T\}$ denotes the training set where T indicates the total number of samples, the optimisation scheme can be given as

$$w^* = \underset{w}{\operatorname{argmin}}\, E(\mathbf{w}, w_0|\mathcal{D}) = \underset{w}{\operatorname{argmin}}\, \frac{1}{T}\sum_{n=1}^{T}(g(\mathbf{x}_n|\mathbf{w}, w_0) - C_n)^2 \tag{2.21}$$

Commonly the solution of the scheme is subjected to iterative optimisation methods, for instance *gradient descend* [190]. Although gradient methods are simple and effective, it is

[17] This figure is inspired by [143]. (Note that there is a mistake of "$g(\mathbf{x}) > 0$ for C_1" and "$g(\mathbf{x}) < 0$ for C_2" in Figure 10.1 in [143], which has been corrected here.)

not the only solution. Many other optimisation algorithms can also be used, for instance simulated anneling or genetic algorithms. In addition, it is because of such learning mechanisms that parameters can be directly calculated without assuming probability densities for any class [143].

Last but not least, it is possible that the above arguments of the linear discriminant model can be generalised to the quadratic discriminant model. Introducing higher-order terms on the basis of Equation 2.18 would bring us the corresponding discriminant function. In summary, considering the nature of our study is a two-class problem, the associated decision boundary of both classifiers are given as

$$\begin{aligned} g_L(\mathbf{x}) &= \sum_{i=1}^{n} w_i x_i + w_0 \\ g_Q(\mathbf{x}) &= \sum_{i,j=1}^{n} w_{ij} x_i x_j + \sum_{i=1}^{n} w_i x_i + w_0 \end{aligned} \tag{2.22}$$

Support Vector Machine with Linear/Quadratic Kernel SVM, as its name may imply, is one of the binary classifiers that utilises support vectors (i.e. an optimised portion of the training data) to *support* the decision boundary for learning tasks. Since it was proposed by Vapnik in 1995 [148], SVM has been a great success in many ML-related applications thanks to its excellent empirical performance. The profound driving force behind is rooted in a simple fact that the number of parameters required to be set in an SVM only relates to the number of training objects, not to the number of features. As a result, it effectively facilities learning situations where the number of features is much larger than the number of training objects, analogous to what we commonly encounter in bioinformatics.

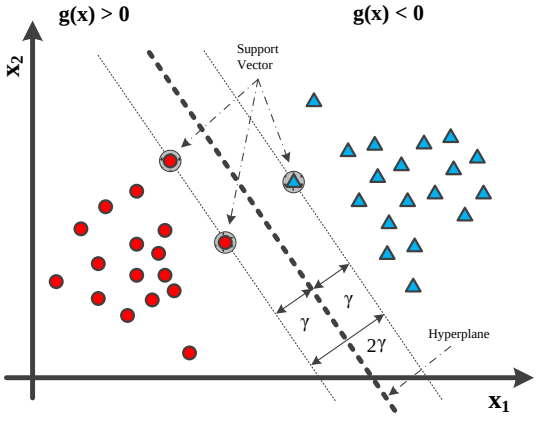


Figure 2.16: An illustrative example of two-dimensional two-class SVM classification.

Here, the theoretical background and the decision boundary of the SVM have to be covered (mainly from [142]). Starting with the basic, a standard SVM uses a linear decision boundary. The discriminant function of the SVM is akin to Equation 2.18. Class labels in this case are different to the previous models, which by default are $\{1, -1\}$ for the SVM. However, the main difference lies in the learning part, as the SVM attempts to maximise the *margin* γ, which is the perpendicular distance to the closest points on either side of the decision boundary (Figure 2.16). The main object in the SVM is straightforward, that one has to maximise the margin; or alternatively double the boundary as 2γ. So given two respective closest points $\mathbf{x}_1$ and $\mathbf{x}_1$ on each side of the hyperplane and the direction perpendicular to the hyperplane $\frac{\mathbf{w}}{\|\mathbf{w}\|}$, we have

$$2\gamma = \frac{1}{\|\mathbf{w}\|}\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) \tag{2.23}$$

Since both the closest points lie on the margin, i.e. $\mathbf{w}^T\mathbf{x} + w_0 = \pm 1$, that leads Equation 2.23 to be

$$\gamma = \frac{1}{\|\mathbf{w}\|} \tag{2.24}$$

Equation 2.24 tells nothing but to maximise $\frac{1}{\|\mathbf{w}\|}$ for margin maximisation. However, constraints for such optimisation have to be complied with. Specifically, $\mathbf{w}$ has to be chosen so that $\mathbf{w}^T x + w_0 = +1$ for all training data in class 1 and $\mathbf{w}^T x + w_0 = -1$ for those in class -1 are satisfied. In effect, it imposes T constraints (recall that T indicates total number of samples in training set $\mathcal{D}$) on the optimisation. To make it easier, maximising Equation 2.24 can be considered as minimising $\frac{1}{2}\|\mathbf{w}\|^2$. Thus, the whole brings us to

$$\underset{\substack{\mathbf{w} \\ \text{s.t. } C_n(\mathbf{w}^T\mathbf{x}_n + w_0) \geq 1, \text{ for all } n}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{w}\|^2 \tag{2.25}$$

Equation 2.25 is basically a constrained optimisation problem. To handle this, T positive Lagrange multipliers are added in order to incorporate the constraints into this objective function. As a result of this, we have

$$\underset{\substack{\mathbf{w},\alpha \\ \text{s.t. } \alpha_n \geq 0, \text{ for all } n}}{\operatorname{argmin}} \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{n=1}^{T} \alpha_n(C_n(\mathbf{w}^T\mathbf{x}_n + w_0) - 1) \tag{2.26}$$

To reach the optimum, partial derivatives of Equation 2.26 with respect to $\mathbf{w}$ and w_0 must be zero. Thus, we have

$$\frac{\partial}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^{T} \alpha_n C_n \mathbf{x}_n \tag{2.27}$$

$$\frac{\partial}{\partial w_0} = -\sum_{n=1}^{T} \alpha_n C_n \tag{2.28}$$

Setting Equation 2.27 to zero gives us the optimum

$$\mathbf{w} = \sum_{n=1}^{T} \alpha_n C_n \mathbf{x}_n, \quad \text{where} \quad \sum_{n=1}^{T} \alpha_n C_n = 0 \tag{2.29}$$

Eventually, substituting Equation 2.29 back into Equation 2.26 leads us to a new objective function that must be maximised with respect to α_n.

$$\begin{aligned} & \tfrac{1}{2}\mathbf{w}^T\mathbf{w} - \textstyle\sum_{n=1}^{T} \alpha_n (C_n(\mathbf{w}^T\mathbf{x}_n + w_0) - 1) \\ = \quad & \textstyle\sum_{n=1}^{T} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{T} \alpha_n \alpha_m C_n C_m \mathbf{x}_n^T \mathbf{x}_m \\ & \text{s.t. } \alpha_n \geq 0, \ \textstyle\sum_{n=1}^{T} \alpha_n C_n = 0, \ \text{for all } n \end{aligned} \tag{2.30}$$

Equation 2.31 is known as the dual optimisation problem, which can be solved by a constrained convex quadratic programming technique [143]. More importantly, since it is convex, a maximum point can be achieved.

Once we solve for α, we can see that among T of them most of $\alpha = 0$ and normally quite a few $\alpha > 0$. For training samples whose $\alpha > 0$, they are called *support vectors*; for those where $\alpha = 0$,they are seen to lie fairly far away from the hyperplane, and therefore have no effect on it. The didactic example has shown the support vectors (SVs) that matter, and the rest carry no information about the hyperplane (or decision boundary). In order to make a prediction on a new sample, we have

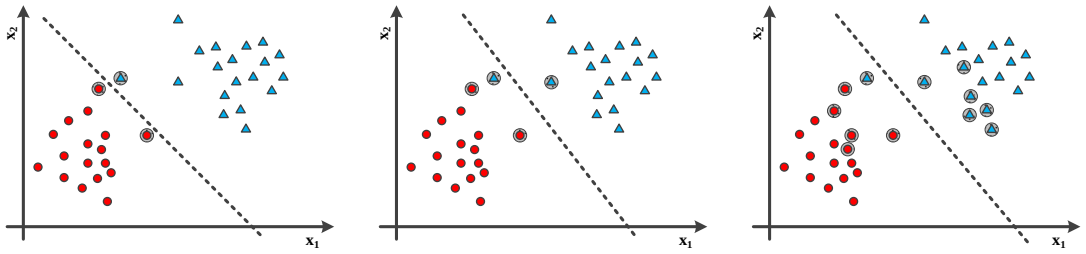
$$\begin{aligned} C &= sgn(g(\mathbf{x})) = sgn(\mathbf{w}^T\mathbf{x} + w_0) \\ \text{where } \mathbf{w}^T &= \sum_{n=1}^{T} \alpha_n C_n \mathbf{x}_n, w_0 = C_l - \mathbf{w}^T\mathbf{x}_l \end{aligned} \tag{2.31}$$

given C_l and x_l correspond to any one of the closest points to the boundary that fits $C_l(\mathbf{w}^T\mathbf{x}_l + w_0) = 1$. With that w_0 can be determined.

So far we have touched on how the SVM is learnt to obtain a hyperplane. Imagine a scenario where the SVM is built with three SVs as in Figure 2.17(a)[18]. The hyperplane

[18]This figure is inspired by Figure 5.16, [142].

in this case seems to be too much constrained by the SVs. That is because the optimisation procedure we just discussed (the constraint shown in Equation 2.25) forces the hyperplane to lie exactly between the two classes so that all training samples lie on the correct side of the boundary, rending itself a *hard margin* SVM. This type of SVM seems harsh, though, and no sensible hyperplane would be possible if the training samples were noisier. Relaxation of the constraints may be a good option, and this essentially leads us to a *soft margin* SVM.



(a) (b) (c)

Figure 2.17: Demonstration of hard margin and soft margin SVM, featuring regularisation parameter C in the latter. (a) hard margin SVM; (b) soft margin SVM when $C = 1$ (c) soft margin SVM when $C = 0.01$.

To realise a soft margin SVM for potential points that lie on the wrong side of the boundary, a slack parameter ξ_n is added to slacken the constraints

$$C_n(\mathbf{w}^T\mathbf{x} + w_0) \geq 1 - \xi_n \tag{2.32}$$

where $\xi_n \geq 0$. That means, if $0 \leq \xi_n \leq 1$, the point $(\mathbf{x}_n, C_n)$ lies on the correct side of the boundary but within the margin; or if $\xi_n > 1$, the point $(\mathbf{x}_n, C_n)$ lies on the wrong side of the boundary. Along with this parameter, we have the regularisation parameter C which works as a penalty factor in the optimisation. Therefore, Equation 2.25 becomes

$$\underset{\substack{\mathbf{w},\alpha \\ \text{s.t. } \xi_n \geq 0 \text{ and } C_n(\mathbf{w}^T\mathbf{x}+w_0) \geq 1-\xi_n, \text{ for all } n}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{n=1}^{T}\xi_n \tag{2.33}$$

With these two new parameters, again similar approach is run through as before and eventually we reach a new dual optimisation problem

$$\sum_{n=1}^{T} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{T} \alpha_n \alpha_m C_n C_m \mathbf{x}_n^T \mathbf{x}_m \tag{2.34}$$
$$\text{s.t.} \quad 0 \leq \alpha_n \geq C \text{ and } \sum_{n=1}^{T} \alpha_n C_n = 0, \text{for all } n$$

With C we effectively control to what degree training points could lie within the margin, or even on the wrong side of the boundary. In other words, the increase of C would generally result in less SVs, with each of them playing higher weights to support the boundary; the decrease of C, however, would lead to more SVs, allowing more chances to correct the potential mistakes. Figure 2.17(b) and 2.17(c) essentially demonstrate the effect of C, with $C = 1$ and $C = 0.01$ respectively. Quite clearly, different C values would bring in different numbers of SVs. So computational complexity taken to learn and make predictions can be an issue in some settings. As we will also see in Chapter 5, C will play the key role in trading off the accuracy and computational complexity of SVM for predictions. Moreover, to find an optimal C for applications is generally suggested under cross validation. From there it can be fine-tuned so as to find the trade-off between margin maximisation and error minimisation.

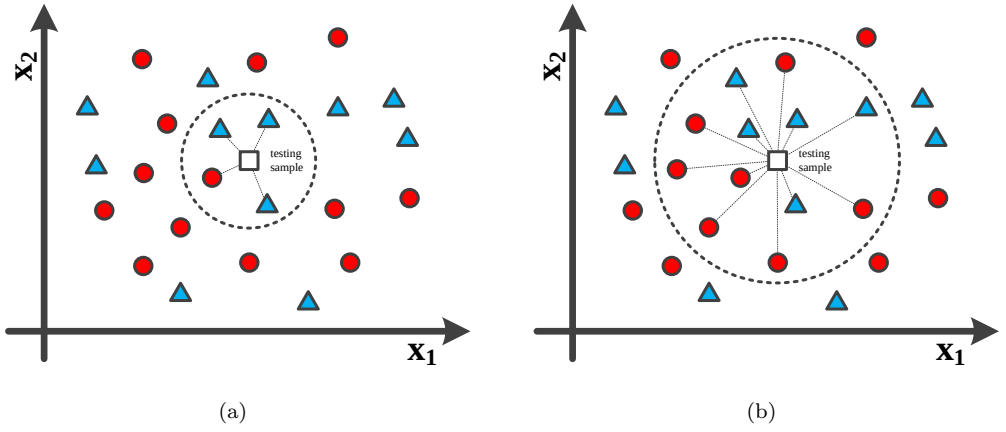
Probably the most important feature of the SVM is the deployment of kernel, a scheme that transforms each data point into a new transformed space. As we can see in Equation 2.35, the inner product of $\mathbf{x}_n^T \mathbf{x}_m$ can actually be viewed as $k(\mathbf{x}_n^T, \mathbf{x}_m)$ in the form of a kernel function. From there transformation from the original space (normally low-dimensional) to a new space (normally high-dimensional) can be done explicitly. The advantage of using a kernel can be greatly manifested in comparison to other learning algorithms, because there exists a kernel function that will allow the data, which are nonseparable in lower dimensional space, to be linearly separable when transforming into higher dimensional space [191]. So in our study, we will be using two types of kernel [192]:

$$\text{linear kernel} \qquad k_L(\mathbf{x}_n^T, \mathbf{x}_m) = \langle \mathbf{x}_n, \mathbf{x}_m \rangle = \mathbf{x}_n^T \mathbf{x}_m \tag{2.35}$$
$$\text{quadratic kernel} \quad k_Q(\mathbf{x}_n^T, \mathbf{x}_m) = \Big(\langle \mathbf{x}_n, \mathbf{x}_m \rangle + 1 \Big)^2 = \Big(\mathbf{x}_n^T \mathbf{x}_m + 1 \Big)^2 \tag{2.36}$$

As a result, according to Equation 2.31 the prediction function can be rewritten as

$$C = sgn\Big(\sum_{n=1}^{T} \alpha_n C_n k(\mathbf{x}_n^T, \mathbf{x}) + w_0 \Big) \tag{2.37}$$

k-Nearest Neighbours (k-NN) k-NN is a nonparametric and nonprobabilistic classifier, as it neither makes assumptions about parametric form of the decision boundary nor estimates any probability of the classes. It simply does one job – finds k training points that are closest to the testing sample and subsequently assigns the majority class amongst these neighbours to this sample. The distance measure between the testing sample and the training ones commonly uses Euclidean distance, while others are also possible (e.g. Mahalanobis distance). Figure 2.18 shows an illustrative example of how k-NN actually works. It can be seen that when k = 4, the major class is a triangle and thus the testing sample is assigned to triangle; but when k = 12, the major class is a circle and thus testing sample is assigned to this class instead. Naturally the choice of k would lead us to another question: how to determine k in specific applications? The general rule is to apply cross validation methods in multiple runs, so that errors generated by various k values can be observed and thus a decision on k with the lowest error can be made.



(a) (b)

Figure 2.18: An illustrative example of two-dimensional two-class k-NN classification. (a) if k=4, a square is assigned to a triangle class; (b) if k=12, a square is assigned to a circle class.

In our study, we use Euclidean distance as the distance measuring method. This leads us to the calculation of distance

$$d(\mathbf{x}_i, \mathbf{x}_n) = \sqrt{\sum_{j=1}^{n} (x_{i,j} - x_{n,j})^2} \tag{2.38}$$

where $x_{n,j}$ indicate the training data at dimension j. In this study, we set k = 3 under trials of cross validation throughout the thesis.

2.3.5 Performance Evaluation

Having discussed the classification algorithms, we should now discuss the ways of evaluating their performance. In this section, the definition and justification of using k-fold Cross Validation (CV) is covered. Besides, raw classification accuracy, sensitivity, specificity as the main three metrics for comparison between classification algorithms in our study are also covered. It is believed that leveraging these metrics would lead us to the optimal classifier throughout different phases of our study. In addition to those above, a more detailed discussion on assessing the performance of classification methods is given in [193]. Having thought through the guidelines introduced in that article, choices of the above metrics are actually made accordingly.

2.3.5.1 k-fold Cross Validation

To validate the predictive performance of a classification model, validation data should be deployed. In fact, validation data could be provided separately or created by removing some data from the original training set. However, the performance of the classification model ought to be varied due to the choice of data in our validation set, and hence inconsistency in the assessment may exist. To overcome this problem, k-fold CV is suggested. It is a technique that takes efficiently full advantage of the data we have at hand, especially when the dataset is small [142]. As it divides the data into k blocks of equal size, every block takes its turn as a validation set while the training set, therefore, consists of the other k-1 blocks, as shown in Figure 2.19. With k times (the folds) the validation set changes, the resulting k performance values are averaged so as to derive the final value for assessment. In addition, a practical study specifically in cross-validation for accuracy estimation and model selection [194] has managed to shown that, 10-fold CV may be better than the more expensive leave-one-out CV (where k equals the total number of samples). Although 10-fold CV is commonly used, it still depends on the available dataset and its number of samples being analysed, where in some cases leave-one-out CV may be a better choice [195].

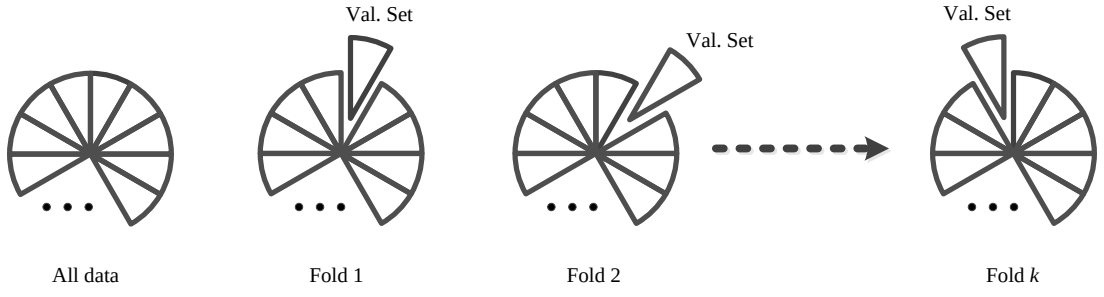


Figure 2.19: k-fold cross validation.

2.3.5.2 Assessment of Classification Algorithm

To assess the performance of classification algorithms, common and indicatively important metrics are used throughout our work. Briefings of the metrics below are mainly extracted from [142].

Raw Classification Accuracy: Raw Classification Accuracy (Acc) is usually used when a measure of performance is required. The loss function, which measures how bad the overall performance of a classifier is, is 0/1 loss in this case. The loss could be 0 or 1, depending on whether the prediction for one particular test point is incorrect or correct. In other words, if the prediction is perfectly accurate, the loss function is zero; otherwise the loss function presents the portion of samples for which the classifier is wrong when averaged over all samples. Therefore, it can be referred to as the error rate. In our case, instead of giving the error rate, the correct rate is used.

$$Acc = 1 - \frac{\text{incorrectly classified samples}}{\text{total number of samples}} = \frac{\text{correctly classified samples}}{\text{total number of samples}} \times 100\% \tag{2.39}$$

One concern about raw classification accuracy is that, when the number of classes is not balanced, attention should be drawn in the usage of this metric. However, as we will see later in this report, a balanced condition has been made to avoid this type of problem.

Sensitivity and Specificity: Mainly used in statistics and medicine, sensitivity and specificity are designated to measure the proportion of actual positives and the proportion of actual negatives which are correctly identified by the classifier, respectively. Before giving the derivation of sensitivity and specificity, four summary metrics[19] should be introduced:

- True Positives (TP) – the number of positive samples that are classified correctly (e.g. diseased people diagnosed as diseased);
- True Negatives (TN) – the number of negative samples that are classified correctly (e.g. healthy people diagnosed as healthy);
- False Positives (FP) – the number of positive samples that are classified wrongly (e.g. healthy people diagnosed as diseased);
- False Negatives (FN) – the number of negative samples that are classified wrongly (e.g. diseased people diagnosed as healthy).

[19] The definition of these metrics are adopted in general, but by no means as a 'golden standard'; should specific cases raise, an appropriate definition is needed to suit specific studies. Examples can be found in [6]

With these metrics in mind, sensitivity and specificity can be given as:

$$Sen = \frac{TP}{TP + FN} \tag{2.40}$$

$$Spe = \frac{TN}{TN + FP} \tag{2.41}$$

2.4 Concluding Remarks

In this chapter, we have broadly reviewed a range of clinical and technical concepts closely related to the thesis. The basic anatomy of the heart, such as heart structure and electrical condition system, as well as the electrophysiology and electrocardiography including the basic components of the ECG and lead systems have been introduced. Signal processing covering digital domain has also been given, along with the preliminaries of wavelet transform and some outlines of ECG processing. Furthermore, a review of machine learning on several aspects has also been presented, including feature construction and feature selection, as well as classification algorithms used in this thesis and evaluation methods of classification performance. In the next chapter, we will start to discuss ECG feature detection, which falls into the category of signal processing and also serves as the input to machine learning algorithms.

Chapter 3

Automatic Feature Detection of ECG Fiducial Points

In Chapter 2, the importance of feature construction has been stressed on. Utilising the domain knowledge would definitely help find out the clinically useful features for ECG classification. So, to perform extraction of local ECG fiducial points (i.e. the time point at which clinically important cardiac event happen) in remote monitoring environment, automatic feature detection algorithm is needed. However, as for the very nature of mobile applications of telemonitoring, it imposes several constraints to the entire system design, namely accuracy, energy consumption, area, etc. Without doubts, these factors motivate people to formulate appropriate algorithms in order to satisfy different system requirements. With this in mind, two ECG feature detection algorithms that aim at different strategies are proposed in this chapter to extract 11 key features from a full ECG heartbeat.

The rest of this chapter is organised as follows: Section 3.1 briefs the motivation behind this chapter. Section 3.2 highlights the basic principle of our first ECG feature detection algorithm – Time-Domain Morphology and Gradient Algorithm (TDMG)[1] and detailed description of the methodology. Similar way is also taken in Section 3.3 to introduce our second ECG feature detection algorithm – Hybrid Feature Detection Algorithm (HFDA)[2]. Section 3.4 covers the validation of our algorithms in comparison with the state-of-the-art algorithm as well as clinical standard. In the end, Section 3.5 concludes this chapter.

[1]This work was a joint effort and has appeared as "A Time-Domain Morphology and Gradient based Algorithm for ECG Feature Extraction" in [117]. Other authors including Evangelo Mazomenos are greatly acknowledged.

[2]This work was a joint effort and has partly appeared as "A Low-Complexity ECG Feature Extraction Algorithm for Mobile Healthcare Applications" in [196]. Other authors including Evangelo Mazomenos are greatly acknowledged.

3.1 Motivation

Having mentioned the design constraints in our context, there are two major factors that mainly play the key role in the design of ECG feature detection algorithms – accuracy and power consumption: the performance of the algorithm is reflected as accuracy while the computational complexity of the algorithm is directly related to power consumption. In most cases, algorithms that are of high accuracy consume energy several order of magnitude larger than the simple ones. However, the performance of these simple algorithms may not be acceptable in practical scenario. Existing algorithms, namely [138, 197, 198] exhibit very complex signal processing methods. Because of the high complexity, they are not suitable to be implemented in an on-body sensor, as otherwise the battery will run out dramatically.

This has essentially prompted us to consider two different strategies for developing ECG feature detection algorithms: (1) High accuracy: the algorithm will be fully designed to achieve high accuracy in ECG feature detection, without much concern of computational complexity. (2) Balance between accuracy and computational complexity: in this case, the algorithm will be designed to utilise simple signal processing algorithms with concern of limited computational complexity to fulfil the ECG feature detection task with clinically acceptable accuracy.

From technical point of view, motivation of the first proposed algorithm of ECG feature detection can be given as follows. ECG signals consist of several constituent waves on the isoelectric line, either in positive or negative direction. These waves, also called deflections, normally show low, high and low frequency patterns in order, making ECG a non-stationary signal. Inspired by this fact, Pan & Tompkins [129] utilised the gradient of the waves as well as extended version of gradient of the signal to detect the QRS complex in a long-term ECG signal. From this work, it has been shown that it is possible to process and interpret the rest of the ECG features with gradient methodology. That effectively leads us to deploy similar philosophy in ECG feature detection task.

On the other hand, the other proposed algorithm of ECG feature detection has to be designed to run efficiently and require extremely low computational complexity. Doing so will directly benefit the battery lifetime from energy consumption perspective. In addition, automated on-body ECG diagnosis in remote monitoring systems is dedicated to generate an *alarm* whenever any abnormality is detected. It essentially implies that low computational complexity may be possible in return for relatively low accuracy when making first screening alarm. This matter of fact leads us to believe that the trade-off between accuracy and computational complexity of our prospective algorithm can be made. Therefore, it is possible for our algorithm to achieve less complex but fairly accurate performance, which ultimately satisfies both clinical and engineering expectations.

3.2 Proposed Time-Domain Morphology and Gradient Algorithm

3.2.1 Introduction

The proposed algorithm is a Time-Domain based Morphology and Gradient algorithm (TDMG). It has been targeted to achieve high-accuracy level. Relatively complex algorithms will be deployed in order to fulfil the signal processing tasks. Furthermore, this algorithm is designed to be suitable for single-lead based, i.e. being independent of lead location, and varied types of heart disease categories[3]. In the following, key points of the algorithm are listed before detailed discussion:

- *Basic fundamental*: the morphological characteristics of the ECG waves in time-domain is utilised for gradient-based analysis to extract the important fiducial points.
- *Pre-processing*: zero-phase digital bandpass filter coupled with moving average smoothing to produce noise-clean ECG signal. In addition, moving slope filter is used to produce gradient sequence of the ECG.
- *Initial extraction*: during the process of QRS complex as well as P and T wave, initial extraction of the fiducial points is performed with predefined searching windows and adaptive threshold policy. This stage will be further amended by the refinement stage.
- *Refinement:* given the results from the initial extraction, refinement is done with separated predefined searching windows and adaptive threshold policy to make the results as accurate as possible.
- *Outcome:* finally, the 11 clinical fiducial points are obtained.

3.2.2 Overview of the Algorithm

Figure 3.1 illustrates the overview block diagram of the proposed algorithm. Essentially, the entire algorithm can be divided into two main processing stages. The first stage includes some basic filterings to remove the noise and artefacts of the ECG, plus further processing of the ECG signals to generate basic time sequences for later operations, for convenience named as Filtered ECG (ECG_{filt}), Gradient ECG (ECG_{grad}) and Feature ECG (ECG_{feat}). In the second stage, the main feature detection operations take place.

[3]A number of abnormal heart conditions/diseases exhibit atypical ECG morphologies, for example, QRS fragmentation, biphasic P and T waves. With this in mind, consideration of atypical ECG morphologies has been taken during the design of the algorithm.

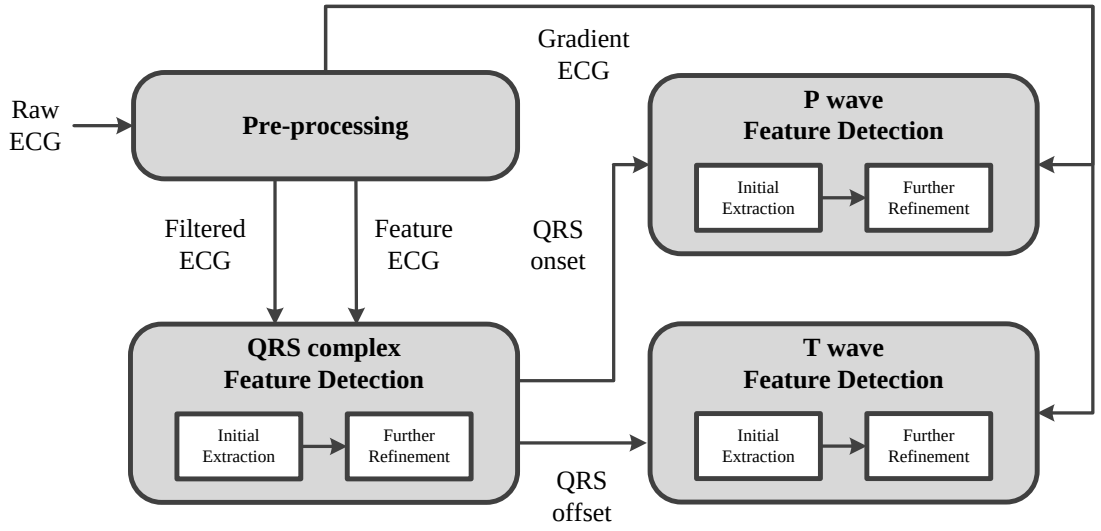


Figure 3.1: The overview block diagram of TDMG.

First, extraction of fiducial points of QRS complex is carried out, followed by the extraction of fiducial points of P wave and T wave. Note that each operation is further refined in order to achieve better detection accuracy as well as potential detection of atypical morphologies.

Since we aim to extract all the required information in time-domain, it is important to understand the morphological representation of the 11 clinical features, such that each can be accurately extracted without being interfered by other features. Keeping this in mind, the basic thoughts behind the algorithm can then be drawn as the following: As for the 11 clinical features, mainly two categorises can be formulated – the peak time points and the onset/offset time points. Essentially, the peak time points are viewed as local extrema around their vicinity. The reason is that in clinical practice with correct placement of the ECG leads, Q and S waves always exhibit as negative waves, while R as positive wave, regardless of the lead location [48]. Whereas for P and T waves, positive deflection or negative deflection can either be seen depending on the location of the lead. Point should also be made that, sometimes biphasic deflection (both positive and negative) may happen in T or P wave, given the presence of certain heart diseases [46]. On the other hand, the onset/offset time points can be obtained by considering the inflection of the start and end of the wave. That means, according to the morphological characteristics of P wave, QRS complex and T wave, it is expected to encounter an abrupt change of the signal amplitude at the start and end point of the wave. Often, such abrupt change would result in a big deflection from the isoelectric line. Therefore, by capturing that change we can determine the onset/offset time points. In the following, each block will be discussed in detail.

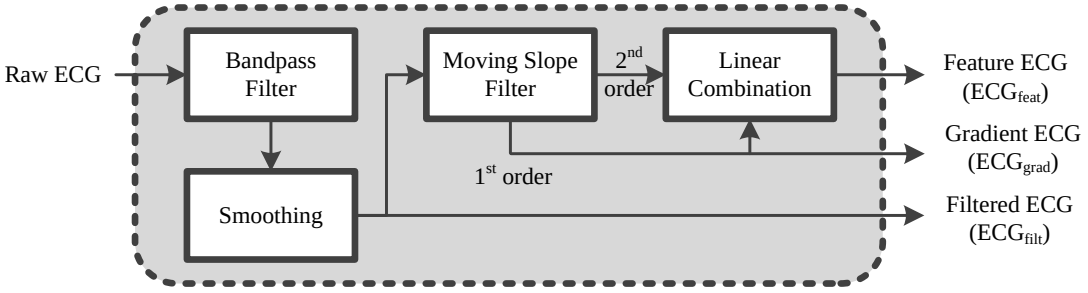


Figure 3.2: The block diagram of pre-processing.

3.2.3 Methodology

3.2.3.1 Pre-Processing Block

As for the pre-processing block shown in Figure 3.2, the main functionality is dedicated to perform initial noise-cleansing operations, and then generate noise and artefact free ECG signal and associated time sequences for further feature detection procedure. Zero-phase Finite Impulse Response (FIR) bandpass filter is deployed here, with cut-off frequencies at 0.5 Hz and 40 Hz. The cut-off frequencies are chosen in accordance with [199]. With such specification, this filter is able to denoise the ECG signal effectively by suppressing baseline drift, Electromyography (EMG) noise, motion artefacts and powerline noise, while retaining the original temporal position of the morphology in the filtered signal thanks to the zero-phase filtering. Following the filtering, a smoothing operation is used here to perform moving average smoothing on the bandpass filtered signal.

At this stage, a noise-free filtered and smoothed ECG signal is available. It is then passed through a moving slope filter. This filter operates on a given sliding window (in this case, 5 samples[4]) and approximates the slope with that window with a first-order model. Thereby, this produces the first-order derivative of the filtered signal, named ECG_{grad}. The reason of obtaining first-order and later second-order derivative is to provide slope information of ECG waves, in order to capture local extrema for peak as well as abrupt change of the ECG wave amplitude for onset and offset. For the ease of later analysis, $ECG_{grad} = |ECG_{grad}|$. In addition, similar to Pan-Tompkins algorithm [129], we also opt to produce a so-called ECG_{feat} signal that is expected to attenuate the P and T wave while enhance the QRS complex. Specifically, this signal is derived from a linear combination of the first and second derivatives (ECG_{grad_2}) of the ECG_{filt} using experimentally verified coefficients, followed by a squaring operation to enhance the characteristics of QRS complex. The associated equation can be given as follows

[4]According to Figure 4 in [129], the frequency response of this filter is almost linear within a range from dc to 40 Hz. In other words, it approximates an ideal derivative over this range.

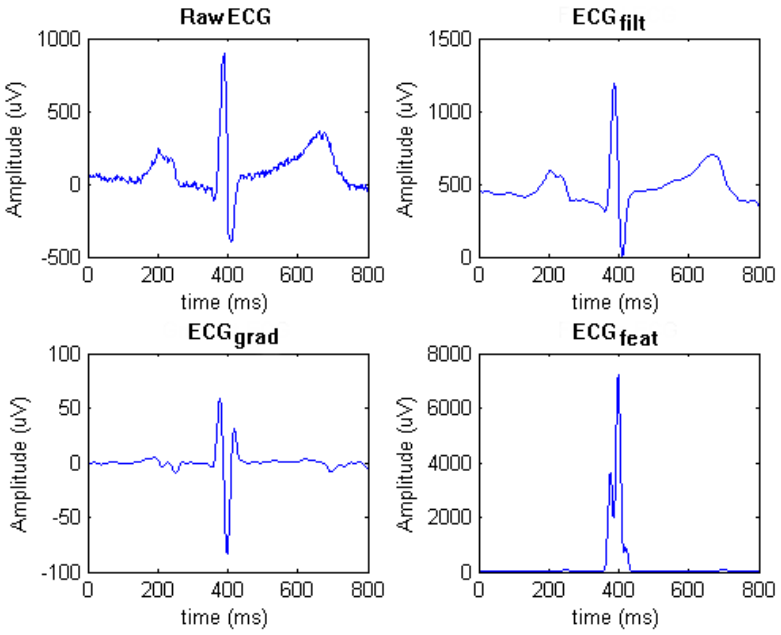


Figure 3.3: The main outcome of the pre-processing stage.

$$ECG_{feat}[n] = (1.3\left|ECG_{grad}[n]\right| + 1.1\left|ECG_{grad\_2}[n]\right|)^2 \tag{3.1}$$

To demonstrate the relation among the raw ECG, ECG_{filt}, ECG_{grad} and ECG_{feat}, Figure 3.3 shows the main outcome of the pre-processing stage. It is fairly obvious that ECG_{feat} manages to enhance the QRS complex to a significant extent compared to the P and T wave. Note that, the novelties of our algorithm with respect to Pan-Tompkins one lie in the fact that (1) Pan-Tompkins uses high-pass and low-pass filter to construct a band-pass filter covering only 5 to 13 Hz, whereas TDMG uses high-order zero-phase FIR bandpass filter and moving average smoothing filter to produce noise-clean ECG signal, with band-pass frequency range 0.5 to 40 Hz; (2) Pan-Tompkins uses 1^{st} order derivative for QRS detection. In TDMG, 2^{nd} order derivative as well as squaring function to produce a more intensified feature ECG signal for R peak detection. With 2^{nd} order instead of 1^{st} order derivative, TDMG has higher chances of finding minor changes in QRS.

3.2.3.2 Fiducial Point Detection

Here, the second stage essentially plays the key role in the entire algorithm. The strategy that we deploy here is three folds. Firstly, in general R peak exhibits a significant positive deflection in the middle of the full ECG heartbeat. Logically enough, it is a good reason to identify R peak as the first instance. Secondly, considering the R peak position as a

reference point, extraction of onset and offset of QRS complex can be executed together with the Q and S peak time points. Eventually, the rest of the features, namely onset, offset and peak of P and T waves can be derived by looking into the ECG segments before and after the detected QRS complex.

Time Frame of the Signal Before we discuss the technical details, discussion on the time frame of the ECG signal should be made. According to the [46], normal duration of a full PQRST complex (worst case) may be about 610 to 620 ms. To assure that our signal processing methods reasonably operate on a ECG signal, the entire heartbeat (from the onset of P wave down to the end of T wave) must lie in the middle of the time frame, plus some isoelectric line segments (100 - 130 ms) included before and after. In our work, the time frame has been generally set to 800 ms. In cases where the heartbeat tends to be longer, compromise of reducing the samples of isoelectric line segments are made.

Window for Temporal Searching It is a concept that will be often encountered in the algorithm. As it will be seen, the estimation of the onset/offset time points for each wave is done by investigating the value of the ECG_{feat} (for QRS complex) or the ECG_{grad} (for P and T wave) within specifically predefined time windows based on certain reference point. During the development of this proposed algorithm, extensive experiments will be carried out to decide the size of the window, which is explained in the following: given a value of threshold (threshold will then be optimised once window frame is fixed), we assign an initial length to the window frame. Here, the default value is 10 ms; (2) depending on the type of the data sequence we are working on (it can be ECG_{filt}, ECG_{grad} and ECG_{feat}) and the fiducial point we are looking for, the length of the window frame has to be adjusted so that it covers sufficient data samples of the sequence to perform the signal processing. For example, the window frame for QRS onset detection in ECG_{feat} was set to be 10 ms long initially, with offset bias 30 ms toward the left of detected R peak. Then the associated signal processing is performed. In this case, it is searching for at least one sample of ECG_{feat} within the window frame that is bigger than the threshold. However, it may so happen that no sample satisfies this condition. To address this, longer window frame is thus required. It is then extended another 10 ms, after which the above procedure is performed again to check whether sufficient samples are obtained. As long as enough samples are observed, the final length of the window frame is thus fixed. Overall, by doing so we can ensure that the time point of the clinical features would fall within the window.

Detection of QRS Complex Within the QRS complex, typically we have R as the prominent positive wave, and Q and S as the negative wave. In standard clinical practice, no matter which lead we look into, the positive deflection always refers to R

wave; similarly, the negative deflection before and after the R wave refer to Q and S wave respectively. In some cases, it may be possible to see only either Q or S wave presents as the most significant feature. This is generally due to the polarity of different leads and the heart electrical axis. For instance, given a normal heart axis, lead aVR would often show prominent Q or S wave. As a result, our single-lead based algorithm has to take this into account.

The two-step process is summarised in Algorithm 1, covering initial extraction (line 2 to line 24) and further refinement (line 25 to line 41) for the characterisation of R peak, QRS onset and offset. In the following, the algorithm will be further discussed in parts, focusing on justification and exemplification of the methods. Note that notation must be introduced beforehand. Taking some examples: 't_{Rpeak}' denotes time instant where peak of R wave locates; 'K_{QRSmax}' denotes the threshold used when finding maximum in QRS sequence; others such as '$sums$' and 'y' refer to temporary internal variables; terms such as 'localmax' and 'max' refer to mathematical operation/function[5]. Same naming convention applies to other variables and the rest of this chapter. Quite a few thresholds are deployed throughout the algorithm, for which more detailed discussion and justification on these thresholds will be covered in Section 3.2.3.3.

1. R Peak Initial Extraction (Line 2 to Line 8): Firstly, a search for local maxima of the entire ECG_{filt} is performed with a threshold K_{QRSmax}. Specifically, the searching routine follows a rule where a point is regarded as a maximum peak if it has the highest value in its vicinity and must be preceded (to the left) by a value lower than K_{QRSmax}. However, it is most likely that the peaks originated from noise may also be captured. Therefore, a further calculation which sums up 30 ms window frame at every maxima point of the ECG_{grad} is executed. By comparing the sums, the maxima that exhibits the largest gradient can be derived, and assigned as R peak. However, the R peak detection might be faulty in situations where the height of R wave is lower than the other extreme points, in particular P peak and T peak . Therefore, it would lead to higher gradient around the P or T peak rather than R peak for which we supposed the highest gradient would have been. As a result, t_{Rpeak} would be incorrect in this scenario.

2. QRS Onset and Offset Initial Extraction (Line 9 to Line 18): To deal with the potentially faulty R peak detection, the next refinement step is carried out. The basic idea is to check whether the R peak has fallen within the QRS complex. However, to do that we first need to obtain the QRS onset and offset. The maximum time point of ECG_{feat} is utilised as reference. Based on this point, estimation of QRS onset and offset can then be performed by comparing the value of the ECG_{feat} with an adaptive threshold K_{QRS} within certain windows (Line 12 and 16). The first and last value in ECG_{feat} that are bigger than the threshold become the

[5]These mathematical operations can be easily found in MATLAB®.

Algorithm 1 Fiducial point detection for QRS complex in TDMG algorithm.

{——— Initial extraction of R peak, QRS onset and offset as well as refinement of R peak ———}
K_{QRSmax} = Eq. 3.3, K_{QRS} = Eq. 3.4
repeat
  $\mathbf{t}_{lm}$ = localmax ECG_{filt}, with preceded value < K_{QRSmax}
  $K_{QRSmax} = K_{QRSmax} \times 0.9$
until num of maxima ≠ 0
$sums_{lm}$ = sums ECG_{grad}(t), $t \in [\mathbf{t}_{lm}$-14ms, $\mathbf{t}_{lm}$+15ms]
t_{Rpeak} = argmax$_t$ $sums_{lm}$, $t \in \{\mathbf{t}_{lm}\}$
$t_{featmax}$ = argmax$_t$ $ECG_{feat}(t)$
$K_{QRS\_1} = K_{QRS\_2} = K_{QRS}$
repeat
  Find t_{QRSon}, the last t that $ECG_{feat}(t) > K_{QRS\_1}$, $t \in [t_{featmax}$-200ms, $t_{featmax}$-30ms]
  $K_{QRS\_1} = K_{QRS} + 0.001 \times Max_{ECGfeat}$
until t_{QRSon} found
repeat
  Find t_{QRSoff}, the first t that $ECG_{feat}(t) > K_{QRS\_2}$, $t \in [t_{featmax}$+30ms, $t_{featmax}$+200ms]
  $K_{QRS\_2} = K_{QRS} + 0.001 \times Max_{ECGfeat}$
until t_{QRSoff} found
if $t_{Rpeak} \in [t_{QRSon}, t_{QRSoff}]$ **then**
  $t_{Rpeak} = t_{Rpeak}$
else
  y_{lm_1} = max $ECG_{filt}(t)$, $t \in [t_{QRSon}, t_{QRSoff}]$, and second highest max y_{lm_2}
  t_{Rpeak} = argmax$_t$ $\{y_{lm\_1}, y_{lm\_2}\}$
end if
{——————— Refinement of QRS onset and offset ———————}
$K_{QRS_3} = K_{QRS}$, $K_{QRS_4} = K_{QRS}$
Find t_{temp} that $ECG_{feat}(t_{temp}) > 20 \times K_{QRS}$, $t \in [t_{QRSon}$-50ms, $t_{QRSon}]$
if $\exists\ t_{temp}$ **then**
  repeat
    Find t_{QRSon}, the last t that $ECG_{feat}(t) > K_{QRS\_3}$, $t \in [t_{featmax}$-200ms, t_{QRSon}-30ms]
    $K_{QRS\_3} = K_{QRS} + 0.001 \times Max_{ECGfeat}$
  until t_{QRSon} found
  t_{Qpeak} = argmin$_t$ ECG_{filt}, $t \in [t_{QRSon}, t_{Rpeak}]$
end if
Find t_{temp} that $ECG_{feat}(t_{temp}) > 20 \times K_{QRS}$, $t \in [t_{QRSoff}, t_{QRSoff}$+100ms]
if $\exists\ t_{temp}$ **then**
  repeat
    Find t_{QRSoff}, the first t that $ECG_{feat}(t) > K_{QRS\_4}$, $t \in [t_{QRSoff}$+30ms, $t_{featmax}$+200ms]
    $K_{QRS\_4} = K_{QRS} + 0.001 \times Max_{ECGfeat}$
  until t_{QRSoff} found
  Find t_{Speak} = argmin$_t$ ECG_{filt}, $t \in [t_{Rpeak}, t_{QRSoff}]$
end if

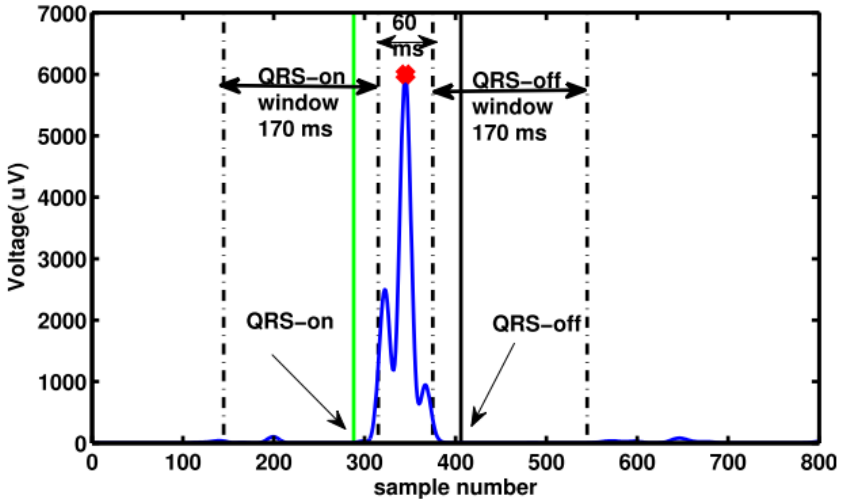


Figure 3.4: QRS boundaries extraction from the feature signal. Green line: QRS onset; black solid line: QRS offset; black dashed line: boundaries of the windows for temporal searching.

onset and offset, respectively. Note that the window frame has been verified after extensive experiments (as mentioned earlier in Window for Temporal Searching). So far, the initial extraction of the onset and offset of QRS is obtained, as shown in Figure 3.4 as an example.

3. R Peak Refinement (Line 19 to Line 24): Having obtained the QRS boundaries, the refinement on t_{Rpeak} is carried out in the following way: now, t_{Rpeak} is checked whether it falls within the QRS boundaries. If it does, it then passes onto the final output of the algorithm, otherwise this initial R peak time point is regarded as faulty. As a result, t_{Rpeak} has to be re-calculated. So, two maxima within the QRS boundaries of ECG_{filt} are obtained. The temporal position of the higher one in amplitude between the two will ultimately be assigned as t_{Rpeak}.

4. Refinement of QRS Onset and Offset (Line 25 to Line 41): Next, a refinement of QRS boundaries is needed, even though we have already attempted to retrieve these features from above. This is because, from physiological and morphological perspective, the QRS complex captured from patients with various heart problems could reveal diversified patterns, for example QRS fragmentation. This type of morphology shows irregular shapes on the QRS waveform, including notching, slurring and slowing [26]. As shown in Figure 3.5[6], notching (a) is defined as a reversal in the QRS trace with an angle ¿ 90^o and amplitude ¿ 0.05 mV; slurring (b) is defined as segment that shows minuscule change in amplitude between successive samples; slowing is defined as a smooth declined rate of change in QRS waveform, with a finite gradient (angle ¡ 90^o) compared to plateau-like slurring.

[6]This figure is excerpted from [26]

From processing perspective, these irregular shapes would lead to valley-like situation in ECG_{feat} within QRS complex. As a result, the initial procedure designed to extract the onset/offset of QRS complex is deemed to be faulty in this case. To overcome this error, refinement is done in the way that it searches for a value of ECG_{feat} that is 20 times higher than K_{QRS} around the vicinity of the initial QRS boundaries (Line 26). The window for temporal searching is located on the left side of QRS onset is set to [t_{QRSon} - 50 ms, t_{QRSon}]. The other window is located on the right side of QRS offset is set to [t_{QRSoff}, t_{QRSoff} + 100 ms]. Having 100 ms on the right side bigger than 50 ms on the left side is due to the fact that we have ST segment after the QRS complex, adding to gradual gradient change of slope: if we have fragmentation at R peak, the next point of ECG_{feat} on the right side that satisfy the threshold is further than the left side one, as ST segment occasionally adds to graduate change of slope and thus makes ECG_{feat} to exhibit values higher than the threshold for a period of time. Now, if the searching manages to find a time point satisfying the threshold (Line 27 and 35), that means the initially estimated onset/offset of QRS in the ECG_{feat} actually lie at the bottom of the valley. It is considered to be erroneous detection. In other words, this implies the true onset/offset of QRS should locate beyond the valley. Therefore, new searches in a smaller window on each side are performed. This time, much accurate location of QRS onset and offset can be derived. As an example, Figure 3.6 shows a particular case of fragmented QS peak. This notch of the QS peak directly results in valley-like situation in ECG_{feat}, which thereby flaws the initial estimation of QRS offset. To tackle this, the proposed refinement method effectively identifies it and makes successful correction. Finally, given the refined R peak and QRS onset and offset, these two peak time points (t_{Qpeak} and t_{Speak}) can be extracted as local minimum between t_{QRSon} and t_{Rpeak} as well as t_{Rpeak} and t_{QRSoff} from ECG_{filt}, respectively.

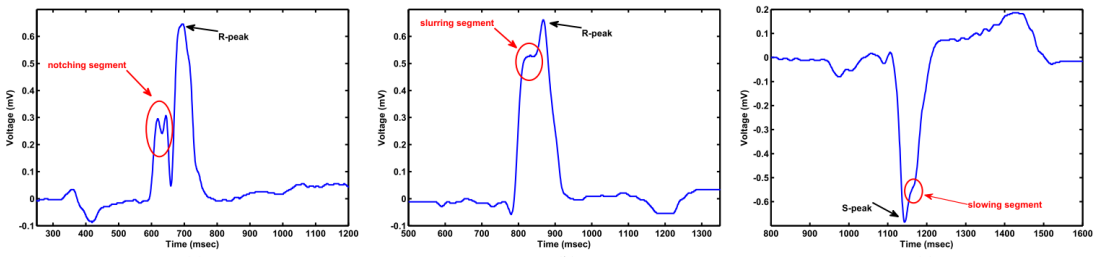
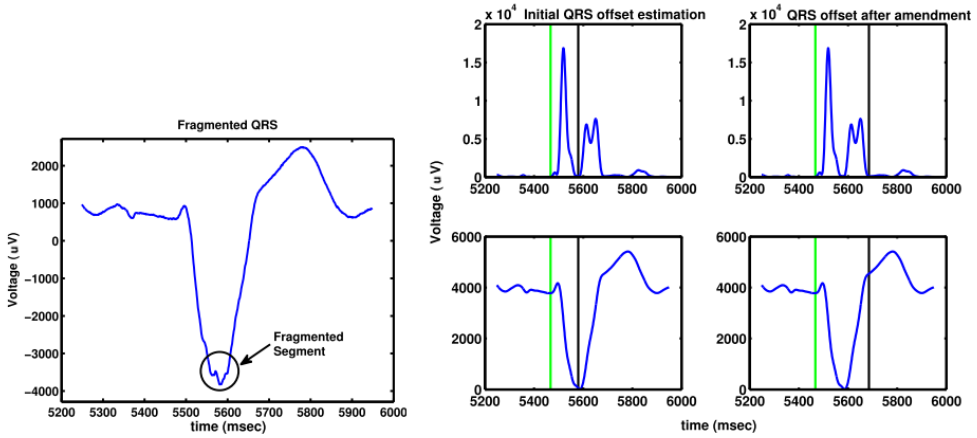


Figure 3.5: Three types of fracmentation: (a) notching; (b) slurring; (c) slowing.

Detection of P and T wave Apart from QRS complex, there are P and T wave present in a heartbeat. They both exhibit either convex or concave wave (relatively low frequency) with respect to isoelectric line. Whether convexity or concavity shows up, it generally depends on either lead position or the heart condition of the patient. As



(a) A fragmented QRS-complex (b) QRS offset before and after the amendment

Figure 3.6: An example of refining the QRS boundaries due to the presence of fragmentation. Green line: QRS onset; black solid line: QRS offset.

our proposed algorithm aims to detect the key features regardless of the lead position, technical details have to be taken into account very carefully.

The process is summarised in Algorithm 2, covering peak extraction (Line 3 to 7) and onset and offset extraction (Line 8 to 14).

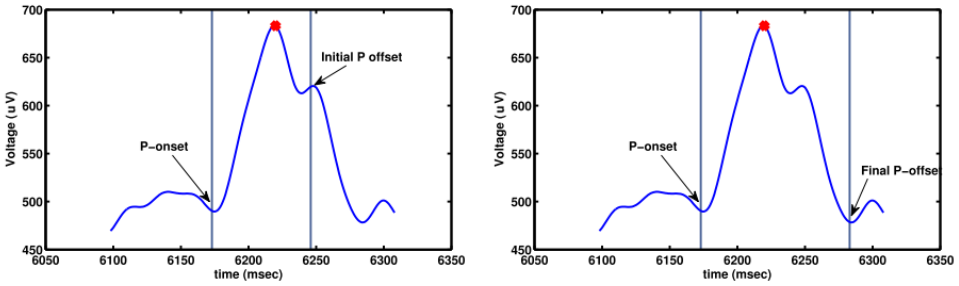
1. P and T Peak Extraction (Line 3 to 7): Unlike R peak detection, P peak is considered to be a local extrema instead of only being the local maxima. This is due to its particularity in either being convex or concave. As a result, local maxima and minima are both extracted with a threshold $K_{PTextrema}$ from the portion of ECG_{filt} that lies before the QRS onset. After that, within the sequences of local maxima and minima, P peak will assign to a particular time point t_{Ppeak} that shows the biggest absolute difference of values with respect to the isoelectric line. Here, the value of the isoelectric line is taken as the value of ECG_{filt} at t_{QRSon}. This is because QRS onset is generally expected to be lying on the baseline.

2. P and T Onset and Offset Initial Extraction (Line 8 to 11): To extract the time point of the onset and offset of P wave, again windows for temporal searching are defined and located before and after the estimated t_{Ppeak}. For P onset, the algorithm searches the last time point of the ECG_{grad} that has a gradient value smaller than a threshold $K_{PTonoff}$. For P offset, the algorithm searches the first time point of the gradient that has value smaller than the same threshold. On the other hand, a fairly similar routine is performed for the extraction of T wave features, except longer windows for temporal searching is required since T wave has a longer duration than P wave.

Algorithm 2 Fiducial point detection for P and T wave in TDMG algorithm.

1. $K_{PTextrema}$ = Eq. 3.5, $K_{PTonoff}$ = Eq. 3.6, $K_{PTamend}$ = Eq. 3.7
   {——————— Peak extraction ———————}
2. **repeat**
3. t_{lmx} = localmax ECG_{filt}, t_{lmn} = localmin ECG_{filt}, with preceded value < $K_{PTextrema}$
   P: $t \in$ [50ms, t_{QRSon}-10ms];
   T: $t \in$ [t_{QRSoff}+20ms, 800ms-500ms]
4. $K_{PTextrema} = K_{PTamend} \times$ 0.9
5. **until** num of extrema $\neq$ 0
6. y_{temp_1} = max $\{ECG_{filt}(t_{lmx}), ECG_{filt}(t_{QRSon})\}$
   y_{temp_2} = max $\{ECG_{filt}(t_{lmn}), ECG_{filt}(t_{QRSon})\}$
7. t = argmax$_t$ $\{y_{temp_1}, y_{temp_2}\}$, P: $t = t_{Ppeak}$; T: $t = t_{Tpeak}$
   {——————— Onset and offset extraction ———————}
8. Find t_{on}, the last t that $ECG_{grad} < K_{PTonoff}$
   P: t_{Pon} and $t \in$ [1ms, t_{Ppeak}-30ms];
   T: t_{Ton} and $t \in$ [t_{QRSoff}+10ms, t_{Tpeak}-50ms]
9. y_{diff} = ECG$_{filt}(t)$ - $ECG_{filt}(t_{QRSon})$, P: $t = t_{Pon}$; T: $t = t_{Ton}$
10. Find t_{off}, the first t that $ECG_{grad} < K_{PTonoff}$
    P: t_{Poff} and $t \in$ [t_{Ppeak}+25ms, t_{QRSon}-15ms];
    T: t_{Toff} and $t \in$ [t_{Tpeak}+45ms, 800ms-10ms]
11. $y_{diff} = ECG_{filt}(t)$ - $ECG_{filt}(t_{QRSon})$, P: $t = t_{Poff}$; T: $t = t_{Toff}$
12. **if** $y_{diff} > K_{PTamend}$ **then**
13. Repeat line 8 to line 11, but with new window frames
    P: [1ms, t_{on}-10ms] for t_{Pon} and [t_{Poff}+10ms, t_{QRSon}-15ms] for t_{Poff}
    T: [t_{QRSoff}+10ms, t_{Ton}-10ms] for t_{Ton} and [t_{Toff}+5ms, 800ms-10ms] for t_{Toff}
14. **end if**

3. P and T Onset and Offset Refinement (Line 12 to 14): In case where P and/or T wave exhibit biphasic (positive and negative deflections happen) or *double humps* (consecutive convex/concave deflections) patterns, extra processing routines must be made, or it may wrongly affect the onset and offset localisation. To do that, values of onset and offset for both P and T wave and value of isoelectric line are compared respectively. If difference bigger than a threshold $K_{PTamend}$ is observed, the onset and offset are deemed to be erroneous. Therefore, an amendment is carried out with the same routine as before, except deploying new windows. Same procedure applies to T onset and offset amendment. An example of performing the amendment routine on P wave with double hump is shown in Figure 3.7. This figure justifies the reason of having such amendment. As we can see, without amendment the initial P offset is located between the humps; however, by applying the amendment P offset can be correctly detected. That means amendment actually helps rectifying the faulty P offset. As an example, Figure 3.8 depicts the illustrative result of detecting the features of P and T wave. As we can see, TDMG searches the P onset and P offset within the windows on both sides, and eventually reach to the points where fairly accurate P onset and P offset should be located. Similarly, T onset and T offset are also located fairly accurately.



(a) Initial P-wave boundaries (b) Amended P-wave boundaries

Figure 3.7: An example of the amendment on P wave with double hump. Black line: P onset and P offset.

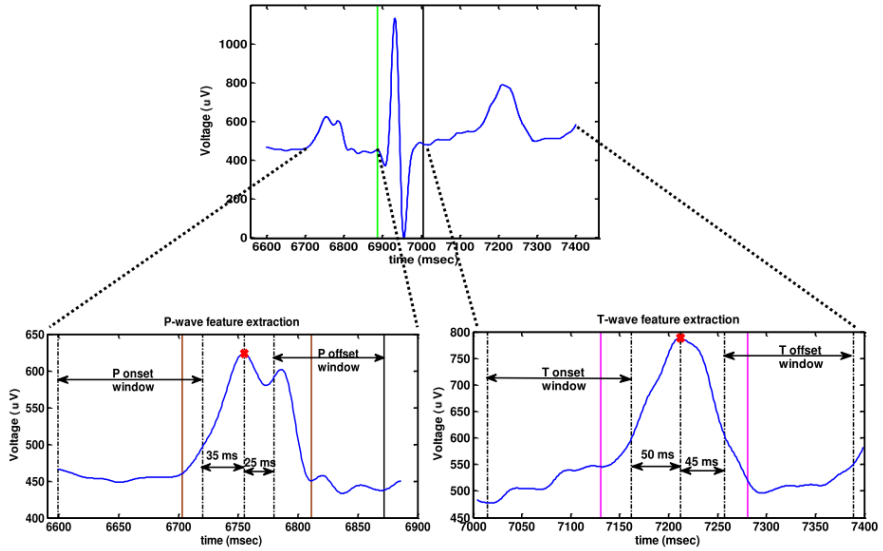


Figure 3.8: Feature detection of P and T wave fiducial points. Green line: QRS onset; black solid line: QRS offset; brown line: P onset and P offset; magenta line: T onset and T offset; black dashed line: window for temporal searching.

3.2.3.3 Justification of the Used Thresholds

In TDMG, it consist of thresholds that are optimally defined. In general, these adaptive thresholds are used to compare with the target feature, so that the detection of fiducial point can be possible. To decide the value of the threshold, it is generally derived in the form of formula that reflects its adaptivity, and obtained by fitting in the required feature(s) of the current ECG signal. In the following, the amplitude thresholds used in this algorithm will be discussed. To do that, they are grouped into wave-oriented types: QRS wave as well as P and T waves. Note that before we proceed to detailed discussion, three basic parameters of amplitude are defined here first.

$$\begin{aligned} Max_{ECGfilt} &= \max\ ECG_{filt},\ t \in [1, 800] \\ Min_{ECGfilt} &= \min\ ECG_{filt},\ t \in [1, 800] \\ Max_{ECGfeat} &= \max\ ECG_{feat},\ t \in [1, 800] \end{aligned} \tag{3.2}$$

Now, for QRS complex the threshold used to decide if a local maxima is found: K_{QRSmax}. This threshold is proportional to the maximum value of the ECG_{filt}, which is given in Equation 3.3; also the threshold used to decide if the QRS onset and offset are found: K_{QRS}, which is given in Equation 3.4. Unlike K_{QRSmax}, this threshold has to adjust itself to different ranges of the maximum value of ECG_{filt}, also with different adaptive factors as well. That is because QRS complex tends to exhibit variety of shapes (could be purely R wave, or QR and RS, or QRS, etc), and therefore it creates varied scenarios in $Max_{ECGfeat}$. Variables like Var_1 in both formula are intentionally set to investigate the optimal values. These variables will be discussed in Table 3.1, to further cover the range of the values of these variables, and more importantly the optimal values that are finally chosen to implement in the algorithm. Same applies to the rest of the variables in this section and Section 3.3.3.2.

$$K_{QRSmax} = Max_{ECGfilt} / (Var_1 \times 10^{\lceil log_{10}(Max_{ECGfilt}) \rceil}) \tag{3.3}$$

$K_{QRS} =$

$$\begin{cases} Var_{2,1} \times Max_{ECGfeat} & \text{if } Max_{ECGfeat} \in [1,\ 200) \quad (3.4a) \\ Var_{2,2} \times Max_{ECGfeat} & \text{if } Max_{ECGfeat} \in [200,\ 850) \quad (3.4b) \\ Var_{2,3} \times Max_{ECGfeat} & \text{if } Max_{ECGfeat} \in [850,\ 1400) \quad (3.4c) \\ Var_{2,4} \times Max_{ECGfeat} & \text{if } Max_{ECGfeat} \in [1400,\ 3000) \quad (3.4d) \\ Var_{2,5} \times Max_{ECGfeat} & \text{if } Max_{ECGfeat} \in [3000,\ 10000) \quad (3.4e) \\ Var_{2,6} \times Max_{ECGfeat} & \text{if } Max_{ECGfeat} \in [10000,\ Inf) \quad (3.4f) \end{cases}$$

For P and T waves, again the threshold used to decide if local maxima/minima is found: $K_{PTextrema}$, which is given in Equation 3.5; the threshold used to decide if onset and offset is found: $K_{PTonoff}$, which is given in Equation 3.6; and the threshold used to judge whether the initial location of onset and offset are erroneous or not, which is given in Equation 3.7. The said three thresholds are all proportional to the difference between maximum and minimum of ECG_{filt}.

$$K_{PTextrema} = (Max_{ECGfilt} - Min_{ECGfilt}) / Var_3 \tag{3.5}$$

$$K_{PTonoff} = (Max_{ECGfilt} - Min_{ECGfilt}) / Var_4 \tag{3.6}$$

$$K_{PTamend} = (Max_{ECGfilt} - Min_{ECGfilt})/Var_5 \tag{3.7}$$

Throughout the algorithm design, the way of formulating the threshold has followed a greedy forward-search approach. That means, locally optimal choice of the threshold, specifically the variables in the formula, are made with the hope of finding a global optimum. Despite it is a standard approach to define threshold in fiducial point detection algorithm [196, 138, 197, 128], it may be argued that these thresholds are heuristically defined and may not be well generalised to other database. To answer this, Table 3.1[7] shows the justification of threshold variables. By analysing the effect of the values within a certain range, we hope to choose a optimal value for specific variable. Coupled with the table, the following tries to explain how the optimal value for a variable is selected on the basis of extensive trials of the values while running the algorithm.

Table 3.1: Threshold variable justification for TDMG algorithm.

Threshold	Variable	Equation	Range	Optimal Value
K_{QRSmax}	Var_1	3.3	[0 1]	0.78
K_{QRS}	$Var_{2,1}$	3.4a	[0 1]	0.1
	$Var_{2,2}$	3.4b	[0 1]	0.015
	$Var_{2,3}$	3.4c	[0 1]	0.009
	$Var_{2,4}$	3.4d	[0 1]	0.003
	$Var_{2,5}$	3.4e	[0 1]	0.01
	$Var_{2,6}$	3.4f	[0 1]	0.0009
$K_{PTextrema}$	Var_3	3.5	[1 10000]	40
$K_{PTonoff}$	Var_4	3.6	[1 10000]	500
$K_{PTamend}$	Var_5	3.7	[1 10000]	15

- Var_1: Given the range of [0 1], towards the lower end the algorithm tends to capture less maxima; but towards the higher end the algorithm captures more. With more maxima it is easier to find the local maxima during the search. Thus finally 0.78 is chosen.

- $Var_{2,1}$ to $Var_{2,6}$: Similarly, within the range of [0 1], towards the lower end there is higher chance of finding a value in ECG_{feat} that is smaller than the threshold; towards the higher end it would be less chance. Therefore, small value is preferred in this case and optimal values are then selected.

- Var_3: Within the range of [1 10000], lower end gives less or even no peaks of P or T wave; and otherwise for the higher end. Since we tend to choose a value that would gather reasonable number of peaks at the first catch, we eventually find the optimal value to be 40.

[7] The range of the variables is defined depending on the lowest and highest value of the associated formula.

- Var_4: Given the range of [1 10000], lower end tends to have more onset or offset temporal point that satisfies the threshold; and otherwise for the higher end. Thus, optimal value is eventually set as 500.
- Var_5: Within the range of [1 10000], towards the lower end difference between the value of onset/offset and isoelectric line would be easier to be observed; and otherwise for higher end. Thus, optimal value is eventually set as 15.

3.3 Proposed Hybrid Feature Detection Algorithm

3.3.1 Introduction

In this section, the proposed algorithm called Hybrid Feature Detection Algorithm (HFDA) will be introduced. By combining DWT and time-domain morphological analysis of the ECG signal, this hybrid algorithm exhibits the benefit of both frequency and time-domain methods. As we have already specified at the beginning of this chapter, this algorithm is designed to be low complex. Also, it is single-lead based and able to tackle biphasic deflection of P and T waves, as what TDMG is capable of doing as well. In the following, key points of the algorithm are listed before detailed discussion:

- *Basic fundamental*: the morphological characteristics of the ECG waves in frequency and time-domain with DWT analysis coupled with gradient-based method is utilised to extract the important fiducial points.
- *Pre-processing*: ECG signal is subjected to DWT analysis, which provides the DWT coefficient sequences for later analysis.
- *Initial extraction of QRS complex*: find out the globe extrema pair from DWT detail coefficient sequence at decomposition level 3 to judge the polarity of the wave deflection for QRS complex. It is then followed by initial extraction of its onset and offset as well as R peak detection with predefined searching windows.
- *Refinement of QRS complex:* time-domain based refinement for QRS onset and offset with adaptive threshold policy is performed. Q and S peak is detected afterwards. Performing this stage helps refine the initial extraction of QRS complex.
- *Extraction of P and T wave*: similarly, DWT detail coefficient sequence at decomposition level 5 is used to extract extrema pair within specific searching windows for judgement of deflection polarity of P and T wave. It is then followed by P and T wave peak detection.
- *Outcome:* ultimately, the 11 clinical fiducial points are obtained.

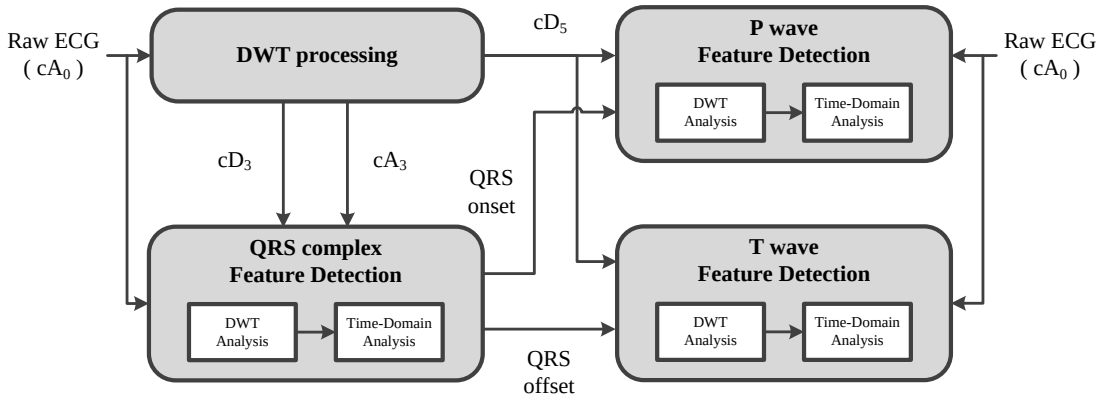


Figure 3.9: The overview block diagram of HFDA.

3.3.2 Overview of the Algorithm

Figure 3.9 illustrates the overview block diagram of the proposed algorithm, where cA and cD represent the approximation and detail DWT coefficients respectively, with suffix meaning the decomposition level of DWT (see Equation 2.11). Similar to the diagram of TDMG in Figure 3.1, the entire process consists of two main processing stages. As the first stage, DWT processing generates the primary DWT coefficient sequences to feed into the subsequence processing blocks. Next, the main feature detection operations take place. DWT based analysis coupled with time-domain based analysis is carried out upon QRS complex to extract its fiducial points. Similar principle is applied to both P and T wave to extract the corresponding fiducial points. In the following sections, each block will be covered in detail.

3.3.3 Methodology

3.3.3.1 Fiducial Point Detection

Time Frame of the Signal Similarly to the case of TDMG in Section 3.2.3.2, the time frame of the signal under investigation is fixed to 800 ms as well. Apart from the reason mentioned in Section 3.2.3.2, DWT in dyadic space also contribute to such decision, as 800 in turn would result in sufficient number of coefficients for analysis when considering higher decomposition levels (e.g. level 5). In addition, we assume P wave, QRS complex and T wave are all present as the constituent ECG waves.

Wavelet Transform and Selection of Mother Wavelet Having covered the basics of Wavelet Transform (WT) in Section 2.2.2, it is not difficult to see the advantages of using WT in biomedical signal processing. In fact, as a well-known technique in this

domain, WT has been extensively used in various researches and applications nowadays. For our HFDA, Discrete Wavelet Transform (DWT) is primarily used. Owing to Multiresolution Analysis (MRA), it can be implemented as filter banks (Figure 2.12) of different frequency ranges. This effectively renders DWT as an effective tool to inherently separate noise and artefacts of the signal. With this in mind, here in our proposed algorithm the dyadic-based DWT is utilised to serve as our basic signal processing tool. To keep the computational complexity low during DWT operations, we opt to use the simplest orthonormal mother wavelet in wavelet family – the Haar wavelet. Its wavelet function $\psi(t)$ and associated scaling function $\phi(t)$ can be found as below. Illustration of Haar wavelet function and scaling function is also given in Figure 3.10.

$$\psi(t) = \begin{cases} +1, 0 \leq t < \frac{1}{2} \\ -1, \frac{1}{2} \leq t < 1 \\ 0, \text{ otherwise} \end{cases} \qquad \phi(t) = \begin{cases} 1, 0 \leq t < 1 \\ 0, \text{otherwise} \end{cases} \tag{3.8}$$

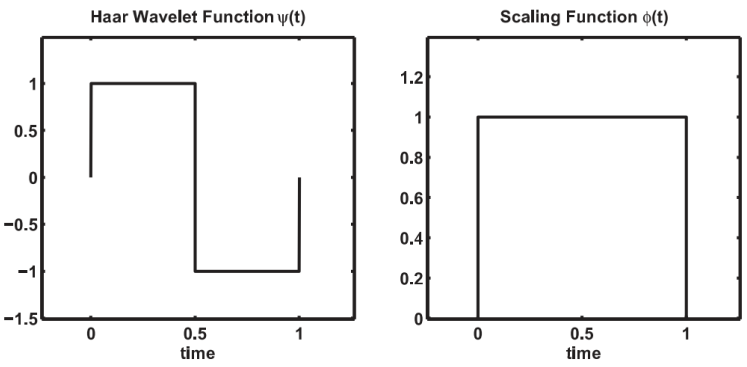


Figure 3.10: Haar wavelet function and scaling function.

Filter coefficients for DWT with Haar [200] can then be given as

$$\begin{aligned} l[0] = l[1] &= \frac{1}{\sqrt{2}} \\ -h[0] = h[1] &= \frac{1}{\sqrt{2}} \end{aligned} \tag{3.9}$$

where $l[n]$ and $h[n]$ stand for low-pass and high-pass filters, respectively.

In the context of DWT-based ECG analysis, Modulus-Maxima Method (MMM)[8] has been effectively used for finding out peaks of characteristic waves at different decomposition level. This method essentially refers to the means of utilising zero crossing point in DWT detail coefficient sequence for peak localisation, as well as the pair of maximum moduli nearby for labelling the peak as positive or negative peak [138]. To demonstrate

[8] A proper example can be found in wave (a) and (b) in Figure 3 in [138].

such method, let us apply MMM on QRS complex with Haar-based DWT as an example. Because the high-pass filter of the Haar transfer function works as a derivative function (Equation 3.9), the output of the filter is able to reflect the *slope* of the input. This in turn translates into the fact that amplitude extrema of the original signal (i.e. cA_0) can be represented as zero-crossing points in detail coefficient at all levels, and maximum slope (either positive or negative slope) of the original signal can be represented as the extrema of the detail coefficients (modulus maxima). Maximum slope refers to the deflection (again, either positive or negative) of the original signal. Bearing these in mind, MMM starts to capture the zero-crossing points for potential QRS peaks as well as the Modulus-Maxima Pair (MMP) for boundaries of QRS complex, both at cD_3 as an example. Here, MMP refers as the pair of two modulus maxima extracted from DWT detail coefficient sequence. The reason why we look for MMP is due to the fact that, the deflection direction of the waves cannot be recognised simply with modulus maxima, unless we look into the pair of them – given a positive deflection of the ECG wave, MMP would appear as a negative minimum followed by a positive maximum; likewise, given a negative deflection, MMP would appear as a positive maximum followed by a negative minimum. In this way, MMM allows to recognise the polarity of the wave under investigation. Examples are given in Figure 3.11. As shown in (a) of this figure, with negative minimum on the left and positive maximum on the right in the WT coefficient sequence we have positive QRS deflection; on contrary, with positive minimum on the left and negative maximum on the right gives us negative QRS deflection.

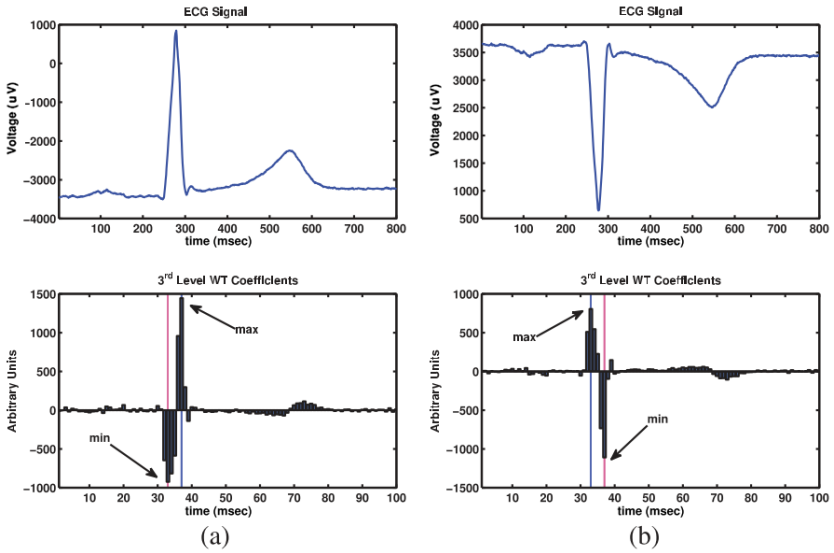


Figure 3.11: Characterizing the direction of the deflection based on the sequence of the WT coefficients extrema pair. (a) Positive deflection. (b) Negative deflection. Magenta line: QRS onset; black line: QRS offset

In comparison to Haar wavelet, quadratic spline wavelet [137, 138] have been in the favour of researchers for ECG analysis – mostly well-known ones include [137, 138]

with use of MMM. Therefore, in this thesis it may be that Haar wavelet might not be appropriate for ECG analysis due to its simplicity. To justify the effectiveness of Haar DWT in terms of highlighting QRS complex as well as P and T waves in time-frequency domain while suppressing noise and baseline wandering, also the effective deployment of MMM for our purpose, representative signals have been selected from PTB database (PTBDB) [201] for explanation.

As shown in Figure 3.12, three ECG signals are present with five decomposition levels of detail coefficients of Haar DWT. Essentially these signals exhibit either baseline wandering, severe noise, or both. From what we can see in the DWT analysis, it is fairly clear that significant noise components mainly lie in the first two decomposition levels. According to what we have said before, it is fairly difficult to obtain the MMP in the vicinity of modulus maxima due to the noise. Therefore, applying MMM on these levels would lead to inaccurate results. Nonetheless, from level 3 onward, high-frequency noises experience suppression to great extent, leaving much cleaner detail coefficients. Such phenomena implies an interesting argument that, it is possible to apply MMM on the sequence of cD_3 (the detail coefficient at level 3) for initial estimation of the QRS features within an entire heartbeat. Further, by just relying on level 3 could be sufficient for QRS feature detection without any consideration of other levels. It can in turn dramatically reduce the computational complexity of the algorithm.

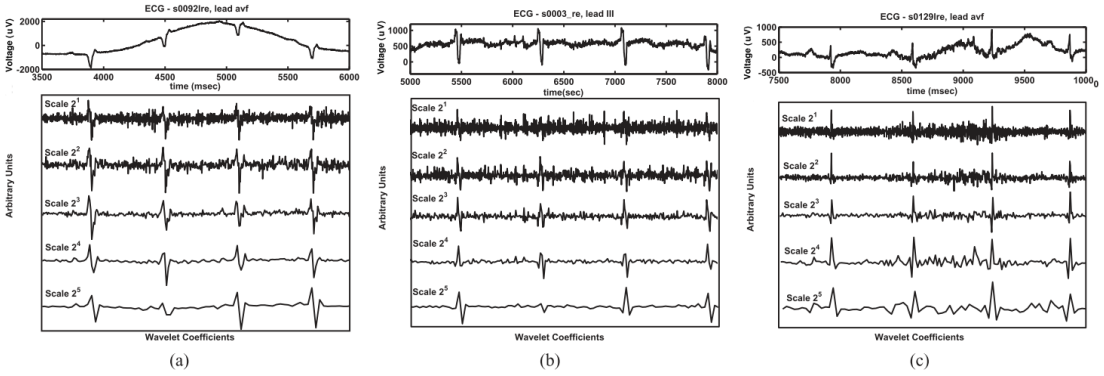


Figure 3.12: Demonstration of Haar DWT analysis in five decomposition levels on ECG samples that exhibits (a) isoelectric line wandering, (b) noise, (c) isoelectric line wandering and noise.

However, it needs to be pointed out that because of the downsampling in DWT, temporal resolution is decreased by 8 folds from initial level down to level 3. There is no doubt that the estimation of QRS features will be affected. Meanwhile, valid frequency components of the QRS may also lie on the other decomposition levels, e.g. level 4. This again may bring in errors in that our approach does not consider other levels except level 3. To tackle these problems while maintaining low complexity of the algorithm, our strategy is to refine the findings of MMM at level 3 by deploying an extra computationally efficient time-domain morphology based approach.

As for the feature extraction of P and T waves, similar approach is taken where only level 5 is considered. There are mainly four reasons for choosing level 5 in this case: firstly, similar to QRS complex, level 1 and 2 are of no option in this case due to noise corruption. Secondly, isoelectric line wandering at level 5 is suppressed to a good extent. Example can be seen in Figure 3.12 (a) and (c), where baseline wandering is effectively filtered out at level 5. Thirdly, P and T wave coefficients at level 5 are deemed to be prominent compared to level 3 and 4. Examples can also be seen in Figure 3.12, in particular (c), where P and T waves can be captured with the use of MMM. Finally, according to the frequency response of the Haar DWT at level 3 and 5 shown in Figure 3.13, it can be observed that the frequency components of P and T waves are lower than QRS complex. It is, therefore, reasonable to choose level 5 for P and T waves analysis.

Detection of QRS complex The detection process is summarised in Algorithm 3, covering initial extraction of QRS onset and offset as well as the R peak finalisation and refinement of QRS onset, offset and peak. Note that, in regards to notation used throughout the algorithm, all follow 3.2.3.2 except now 'n' denotes discrete time instant.

1. DWT (Line 1): First of all, the entire ECG PQRST complex is subjected to Haar DWT. As we have already discussed, up to 5 decomposition levels in dyadic space are considered. To execute the multiscale DWT decomposition, Mallats Algorithm [102] is directly deployed here, where a cascaded filter-bank with high-pass and low-pass filters following Equation 2.11 and 3.9 is structured. In details, cA_0 works as an initial input sequence, passing through the filter-bank. It is then followed by recursive combinations of high-pass and low-pass filter from level to level. High-pass filter produces the detail coefficient cD_j at level j, while low-pass filter generates the approximate coefficient cA_j at the same level. As we

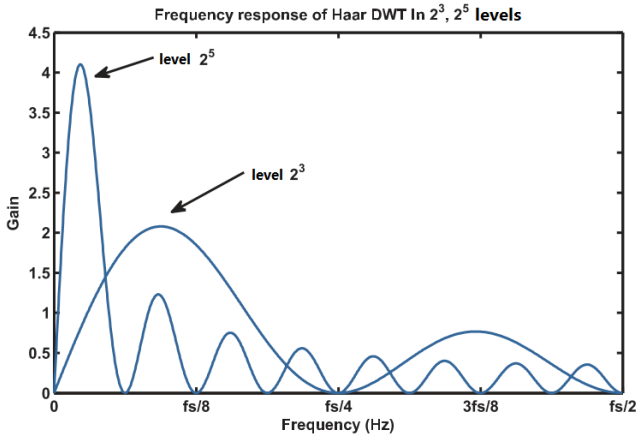


Figure 3.13: Frequency response of the Haar DWT at level 3 and 5, where f_S is the sampling frequency.

Algorithm 3 Fiducial point detection for QRS complex in HFDA algorithm.

1. Apply DWT on raw ECG signal (cA_0, thus cA_3, cD_3 and cD_5 are obtained)
   {———————— Onset and offset initial extraction ————————}
2. n_{cD3GEP_min} = argmin_n cD_3
3. n_{cD3GEP_max} = argmax_n cD_3
4. **if** $n_{cD3GEP\_min} < n_{cD3GEP\_max}$ **then**
5. $n_{cD3GEP\_1} = n_{cD3GEP\_min}$, $n_{cD3GEP\_2} = n_{cD3GEP\_max}$
6. n_{QRSon} = argmax_n cD_3, $n \in [n_{cD3GEP\_1}\text{-}4, n_{cD3GEP\_1}]$
7. n_{QRSoff} = argmin_n cD_3, $n \in [n_{cD3GEP\_2}, n_{cD3GEP\_2}+4]$
8. **else**
9. $n_{cD3GEP\_1} = n_{cD3GEP\_max}$, $n_{cD3GEP\_2} = n_{cD3GEP\_min}$
10. n_{QRSon} = argmin_n cD_3, $n \in [n_{cD3GEP\_1}\text{-}4, n_{cD3GEP\_1}]$
11. n_{QRSoff} = argmax_n cD_3, $n \in [n_{cD3GEP\_2}, n_{cD3GEP\_2}+4]$
12. **end if**
    {———— R peak finalisation, QRS onset, offset and peak refinement ————}
13. **if** $n_{cD3GEP\_min} < n_{cD3GEP\_max}$ **then**
14. n_{Rpeak} = argmax_n cA_0, $n \in [n_{cD3GEP\_min} \times 2^3, n_{cD3GEP\_max} \times 2^3]$
15. **else**
16. n_{cD3NMP_1} = argmin_n cD_3, $n \in [n_{cD3GEP\_1}\text{-}15, n_{cD3GEP\_1}]$
17. n_{cD3NMP_2} = argmax_n cD_3, $n \in [n_{cD3GEP\_2}, n_{cD3GEP\_2}+10]$
18. n_{Rpeak_1} = argmax_n cA_0, $n \in [n_{cD3NMP\_1} \times 2^3, n_{cD3GEP\_1} \times 2^3]$
19. n_{Rpeak_2} = argmax_n cA_0, $n \in [n_{cD3GEP\_2} \times 2^3, n_{cD3NMP\_2} \times 2^3]$
20. n_{Rpeak} = argmax_n $\{n_{Rpeak\_1}, n_{Rpeak\_2}\}$
21. **end if**
22. $n_{Rpeak\_L3} = n_{Rpeak}$ / 2^3
23. Set $K_{QRS\_L3\_1}$ and $K_{QRS\_L3\_2}$ according to Table 3.2
24. cA_{3QRS_der} = cA_3[n] - cA_3[n-1], $n \in [n_{QRSon}\text{-}8, n_{QRSoff}+15]$
25. Val_{diff} = max cA_0 - min cA_0
26. $length_{3QRS\_der}$ = length cA_{3QRS_der}
27. Find n_{QRSon_L3}, the first n that cA_{3QRS_der}[n] > $K_{QRS\_L3\_1}$, $n \in [1, n_{Rpeak\_L3}]$
28. Find n_{QRSoff_L3}, the last n that cA_{3QRS_der}[n] > $K_{QRS\_L3\_2}$, $n \in [n_{Rpeak\_L3}, length_{3QRS\_der}]$
29. $n_{QRSon} = n_{QRSon\_L3} \times 2^3$
30. $n_{QRSoff} = n_{QRSoff\_L3} \times 2^3$
31. n_{Qpeak} = argmin_n cA_0, $n \in [n_{QRSon}, n_{Rpeak}]$
32. n_{Speak} = argmin_n cA_0, $n \in [n_{Rpeak}, n_{QRSoff}]$

mainly concentrate on the cD_3 and cD_5, the rest of the coefficient sequences can be neglected.

2. Initial Extraction (Line 2 to Line 12): Based on what we discussed previously, by using MMP our MMM is able to obtain the temporal position of the deflection, which is expected to have the highest separation from the isoelectric line in the original signal. This is done by calculating the temporal positions of the Global Extrema Pair (GEP) in the cD_3 coefficient sequence, which are n_{cD3GEP_1} and n_{cD3GEP_2}. However, this deflection may correspond to either R peak or Q/S peak. To make the final decision, time-domain based refinement is required and will be discussed later. Now, before we apply the time-domain based refinement,

first approximation of QRS boundaries can be derived by searching in the vicinity of the GEP. The temporal location of QRS onset and offset can then be obtained as the preceding/succeeding extrema (either minimum or maximum) at cD_3 within searching windows (Line 4 to Line 12). Figure 3.14 illustrate an example where this procedure is followed to identify the QRS boundaries. In this particular example, the QRS onset and offset can be found as local maximum and local minimum within respective search window.

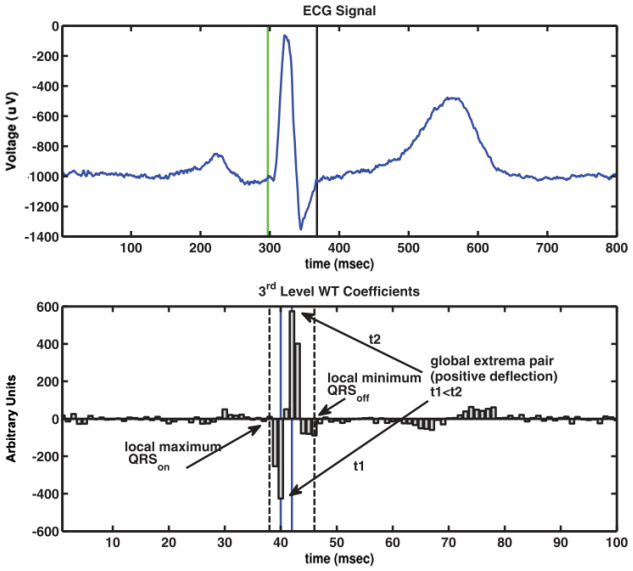


Figure 3.14: MMM for the extraction of the QRS boundaries. The global extrema pair localises the main deflection while the extreme in its vicinity indicate the temporal position of the QRS boundaries. Dashed line: QRS onset and QRS offset.

With steps above, we manage to roughly estimate the peak of the main deflection (either R peak or Q/S peak), and the QRS boundaries. However, due to the down-sampling in DWT architecture, the temporal resolution at level 3 is jeopardised and only one eighth of the resolution of the original signal is available at this level. This, in turn, seriously influences the estimation of the above features. Example can be referred to Figure 3.15, where R peak and QRS offset are not correctly located. As no extra processing is planned to improve the temporal localisation accuracy by doing multi decomposition levels operations in DWT, we opt to perform time-domain based refinement, with belief of achieving better accuracy while maintaining the notion of low complexity.

3. R Peak Finalisation (Line 13 to Line 21): First and foremost, R peak has to be finalised in that it will be considered as a reference point for later refinement of the rest of the features. Given the fact that we have already obtained n_{cD3GEP_1} and n_{cD3GEP_2}, the polarity of the associated deflection is checked. If the deflection is characterised as positive, it is then interpreted as an R wave. In this case, the

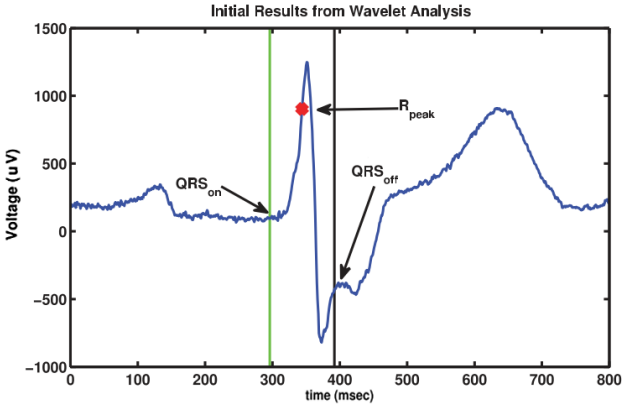


Figure 3.15: Example of less accurate estimation of R peak and QRS offset due to the decreased temporal resolution of DWT at level 3. Green line: QRS onset; black line: QRS offset.

refined R peak time point n_{Rpeak} can be therefore derived directly as the maximum at the original timescale within a time window, which is defined as $[n_{cD3GEP\_1} \times 2^3, n_{cD3GEP\_2} \times 2^3]$ by projecting $[n_{cD3GEP\_1}, n_{cD3GEP\_2}]$ into the original timescale (Line 14). Once n_{Rpeak} is obtained, its corresponding temporal position at level 3 can also be obtained as n_{Rpeak_L3}.

On the other hand, if the deflection is characterised as negative, then it may be interpreted as Q or S peak. But the way to capture Q or S peak will be discussed later in the end of this sub section. Now, to find out n_{Rpeak} given negative deflection is present, we have to further search for minimum/maximum around the global maxima pair so that more potential MMPs can be found and analysed. Hence, we search before n_{cD3GEP_1} for a minimum and after n_{cD3GEP_2} for a maximum. From these two searches, two new MMPs can be formed and thus two new searching windows for finalising R peak can also be formed (Line 16 and 17). Next, we project the two windows from level 3 to original timescale and obtain the maximum value in each of them (Line 18 and 19). Ultimately, the final R peak is picked as the higher one between them (Line 20).

4. QRS Onset, Offset and Peak Refinement (Line 22 to Line 32): In parallel to the R peak time point refinement, QRS refinement is also executed. Similar to the approach we followed in TDMG for QRS boundary detection, we utilise the concept of gradient (derivative) signal here again. Firstly, we select to have approximate coefficient sequence cA_3 for our gradient analysis. This is due to the fact that, cA_3 exhibits a noise-free characteristics after the first two levels of filter banks in DWT. Based on the initially estimated n_{QRSon} and n_{QRSoff}, a dedicated expanded portion of QRS is isolated from the cA_3. Doing so is to ensure that the analysis can run on a clean and complete QRS complex. With this in mind, the process is done by expanding the initial QRS onset to the left by 64 ms (equivalently 8

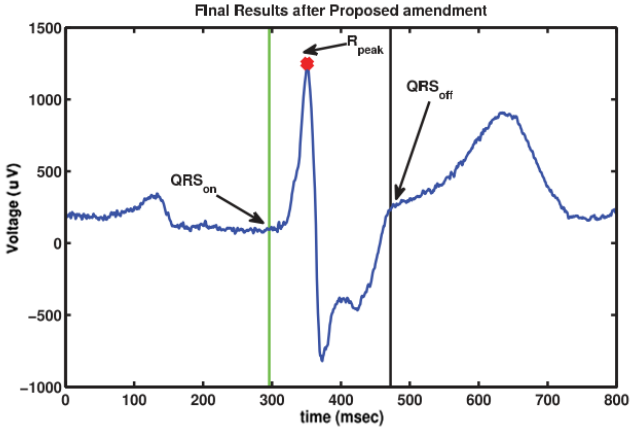


Figure 3.16: QRS final estimation on the signal of Figure 3.15 after the time-domain based refinement. Green line: QRS onset; black line: QRS offset.

samples in cD_3), and the initial QRS offset to the right by 120 ms (equivalently 15 samples in cD_3). On this basis, an approximation of the derivative cA_{3QRS_der} is derived as the backward difference of cA_3 (Line 24). Finally, to finalise n_{QRSon} and n_{QRSoff}, again polarity of the deflection of the global extrema pair has to be taken into account. A threshold policy (with $K_{QRS\_L3\_1}$ and $K_{QRS\_L3\_2}$) also operates upon cA_{3QRS_der}.

Starting with the case of positive deflection, a search routine is made to find out the first time point of cA_{3QRS_der}, where its value becomes higher than a predefined threshold $K_{QRS\_L3\_1}$. Once found, it is assigned to n_{QRSon_L3} (Line 27). Likewise, similar routine is made to find the last time point against threshold $K_{QRS\_L3\_2}$, and ultimately n_{QRSoff_L3} is refined (Line 28). On the other hand, in the case of negative deflection it is fairly similar to the positive one in terms of searching strategy, except the deployment of thresholds. The temporal position of these two in original timescale can be obtained by projecting the position at DWT decomposition level 3 onto original timescale by multiplying 8. An example of deploying the time-domain based refinement to improve the initial estimation (Figure 3.15) is depicted in Figure 3.16.

The final step of the QRS refinement includes the detection of the Q peak and S peak (Line 31 and 32). The temporal location of these two is at the original time scale, where the ECG signal demonstrates minimum value between n_{QRSon} and n_{Rpeak}, and n_{Rpeak} and n_{QRSoff}, for Q peak and S peak, respectively.

Detection of P and T wave Apart from QRS complex, P and T wave are also subjected to detection procedure. Similar to the initial estimation of QRS features, MMM is applied on cD_5 at the portion of the signal that lies before and after detected QRS complex separately. By doing so, we can extract the features of P and T waves. The

process is summarised in Algorithm 4. As we know, these two waves may exhibit either convexity or concavity depending on the ECG lead location or heart disease categories. Again, MMM allows us to recognise them as onset and offset of the wave according to the MMP captured during the process (Line 5, 6, and Line 9, 10). Temporal point of the wave peak can then be detected as time point in the original signal, given that concavity is found (Line 7); or minimum time point, given that convexity is found instead (Line 11). Both should operate within the wave boundaries defined by the modulus-maxima. Example of the results can be found in Figure 3.17, where P onset, P offset, T onset and T offset are relatively correctly located.

Algorithm 4 Fiducial point detection for P and T wave in HFDA algorithm.

1. $length_{cD5}$ = length cD_5
2. P: $n_{QRSon\_L5} = n_{QRSon}$ / 2^5
   T: $n_{QRSoff\_L5} = n_{QRSoff}$ / 2^5
3. n_{cD5MMP_1} = argmin_n cD_5, n_{cD5MMP_2} = argmax_n cD_5
   P: $n \in [1, n_{QRSon\_L5}]$
   T: $n \in [n_{QRSoff\_L5}, length_{cD5}]$
4. **if** $n_{cD5MMP\_1} < n_{cD5MMP\_2}$ **then**
5. P: $n_{Pon} = n_{cD5MMP\_1} \times 2^5$
   T: $n_{Ton} = n_{cD5MMP\_1} \times 2^5$
6. P: $n_{Poff} = n_{cD5MMP\_2} \times 2^5$
   T: $n_{Toff} = n_{cD5MMP\_2} \times 2^5$
7. P: $n_{Ppeak} = \text{argmax}_n$ cA_0, $n \in [n_{Pon}, n_{Poff}]$
   T: $n_{Tpeak} = \text{argmax}_n$ cA_0, $n \in [n_{Ton}, n_{Toff}]$
8. **else**
9. P: $n_{Pon} = n_{cD5MMP\_2} \times 2^5$
   T: $n_{Ton} = n_{cD5MMP\_2} \times 2^5$
10. P: $n_{Poff} = n_{cD5MMP\_1} \times 2^5$
    T: $n_{Toff} = n_{cD5MMP\_1} \times 2^5$
11. P: $n_{Ppeak} = \text{argmin}_n$ cA_0, $n \in [n_{Pon}, n_{Poff}]$
    T: $n_{Tpeak} = \text{argmin}_n$ cA_0, $n \in [n_{Ton}, n_{Toff}]$
12. **end if**

3.3.3.2 Justification of the Used Thresholds

Similarly to Section 3.2.3.3, this section discusses the amplitude thresholds used in HFDA algorithm. Because no adaptive threshold policy is used for P and T wave, we only focus on the threshold for QRS wave. Here, the thresholds are defined adaptively according to the polarity of the deflection and the signal amplitude. Note that, the values of the thresholds are set as a dyadic fraction of amplitude difference (Val_{diff}) of maximum and minimum of the raw ECG, so that the searching routine could adapt itself correctly to varied signals. All have been summarised in Table 3.2.

In Table 3.2, variables are also set to search for the optimal value. To do that, Table 3.3 shows the range of the search for each variable and also the optimal value. Because of

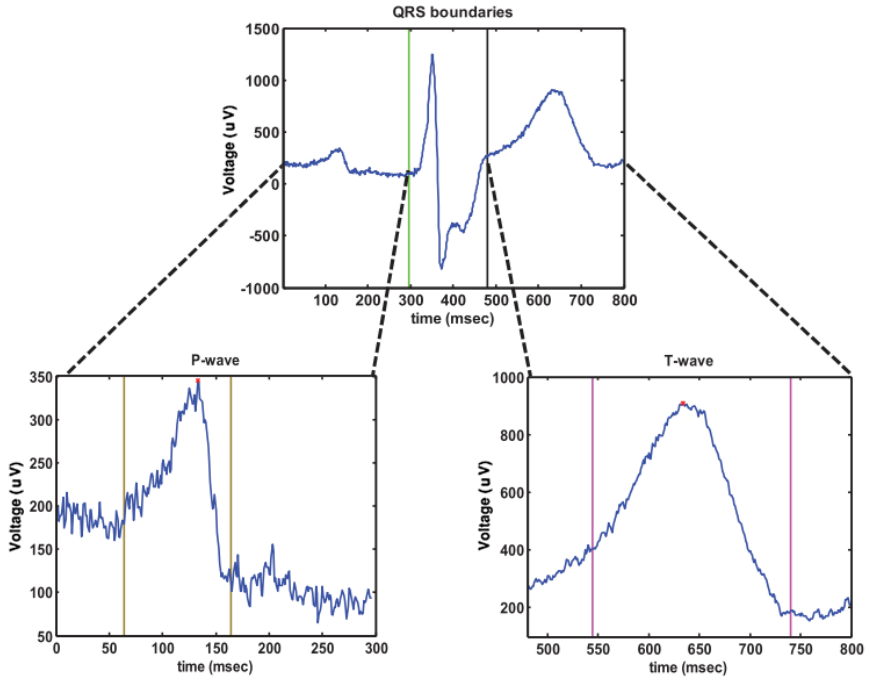


Figure 3.17: P and T wave feature detection. Green line: QRS onset; black line: QRS offset; brown line: P onset and P offset; magenta line: T onset and T offset.

Table 3.2: Definition of adaptive thresholds for refining the QRS onset and offset in HFDA.

Val_{diff}	**Positive Deflection**		**Negative Deflection**	
	$K_{QRS\_L3\_1}$	$K_{QRS\_L3\_2}$	$K_{QRS\_L3\_1}$	$K_{QRS\_L3\_2}$
[1 2000)	$Val_{diff}/Var_{6,1}$	$Val_{diff}/Var_{8,1}$	$Val_{diff}/Var_{7,1}$	$Val_{diff}/Var_{9,1}$
[2000 4000)	$Val_{diff}/Var_{6,2}$	$Val_{diff}/Var_{8,2}$	$Val_{diff}/Var_{7,2}$	$Val_{diff}/Var_{9,2}$
[4000 Inf.)	$Val_{diff}/Var_{6,3}$	$Val_{diff}/Var_{8,3}$	$Val_{diff}/Var_{7,3}$	$Val_{diff}/Var_{9,3}$

the similarity of the variable attributes, we choose to analyse them as a whole instead of individual discussion. Given the range of [1 4000], it is not difficult to see that, towards the lower end of the range it means higher threshold for searching the onset and offset during the QRS refinement, and otherwise for higher end. That effectively means, to refine the QRS boundary to more accurate level, higher threshold is preferred. Despite so, we cannot simply use one threshold and apply it to different scenarios as in 3.2. Depending on the polarity of the deflection and the range of Val_{diff}, certainly different optimal values have to be adopted. Ultimately, the optimal values are chosen as shown in Table 3.3.

Table 3.3: Threshold variable justification for HFDA algorithm.

Threshold	Variable	Range	Optimal Value
$K_{QRS\_L3\_1}$	$Var_{6,1}$	[1 4000]	64
	$Var_{6,2}$	[1 4000]	8
	$Var_{6,3}$	[1 4000]	4
	$Var_{7,1}$	[1 4000]	32
	$Var_{7,2}$	[1 4000]	64
	$Var_{7,3}$	[1 4000]	128
$K_{QRS\_L3\_2}$	$Var_{8,1}$	[1 4000]	64
	$Var_{8,2}$	[1 4000]	64
	$Var_{8,3}$	[1 4000]	64
	$Var_{9,1}$	[1 4000]	8
	$Var_{9,2}$	[1 4000]	8
	$Var_{9,3}$	[1 4000]	16

3.4 Validation and Comparison

3.4.1 Validation

In order to validate the effectiveness of the proposed two algorithms, assessment of the performance in terms of accuracy should be made. As there is no golden rule to determine the peak, onset and offset of the ECG waves, the validation of the ECG feature detection algorithms must be done using databases with manual annotations. The basic idea here is to compare manually annotated results of the clinical features of a specific heartbeat to the ones generated by the algorithms, therefore resulting in mean and standard deviation of the error in total. In this section, two well-accepted databases are used, namely QT database (QTDB) and PTB database (PTBDB). Both are publicly available at Physionet [201].

3.4.1.1 ECG Database

In regards to the ECG database we will be using, there are two we have selected from PhysioNet [201]: QTDB and PTBDB. For QTDB, it consists of 105 two-channel ECG Holter recordings, with particular emphasis on including a broad variety of QRS and ST-T morphologies. With each record, cardiologists identified the onset of P wave and QRS complex, peak of P and T wave, as well as offset of P, T waves and QRS complex. The signals were sampled at 250 Hz and last 15 min duration. According to [202], the first 10 minutes are used for algorithm learning and only the last 5 minutes are forwarded to expert cardiologists for manual annotations, from which between 30 and 100 representative beats were annotated. That ultimately leads to 3622 annotated heatbeats. On the other hand, PTBDB contains 549 records from 290 subjects, with

all 15-lead ECG signals sampled simultaneously at 1 KHz for more than 1 min duration. They cover 8 diagnostic classes including myocardial infarction, cardiomyopathy, bundle branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease, myocarditis and health controls (miscellaneous is not included in this case). Note that the reason of using PTBDB in addition to extensively used QTDB is to evaluate our proposed algorithms on standard 12-lead ECG and varies disease categories to demonstrate their effectiveness.

Having introduced both databases, we have to choose records from them and apply our algorithms on heartbeats in a reasonable manner. From QTDB, out of 105 records 27[9] are chosen here, which are those that come with full annotations for every ECG waves (P, QRS and T). Thanks to the manual annotations, we manage to partition the long-term ECG signals into individual heartbeats. Following, both algorithms are applied on each one of them, resulting into two sets of feature detection results on each of the ECG channel. By then the mean and standard deviation of the error between manual annotations and the automatic results can be derived. In addition, to make our results comparable to the state-of-the-art feature detection method, we opt to follow the guideline provided in [138], from which we know only the channel that exhibited less error should be considered in comparison. To conclude the overall mean and standard deviation, the average mean and standard deviation of the error from the 27 records are achieved.

As for PTBDB, 30 subjects with their first record in the database are chosen, covering each of the disease categories available. Since there is no manual annotation from this database, we have to organise expert cardiologist [203] to help annotate the heartbeats. Limited by the man power and time, we chose to pick up one heartbeat from each lead (simultaneously recorded) of each subject for the expert to annotate with help of a graphical annotator interface developed in MATLAB®. This effort resulted in totally 450 heartbeats fully annotated. Note that, due to the absence of wave or poor sampling, manual annotation on that potential wave may not be available. In this case, the relevant results from both algorithms are decided to be neglected.

In addition, as we have 15 simultaneously recorded leads in this database, it may be more appropriate if we extract one global temporal position for each feature from manual annotations and feature detection results of the 15 leads separately. To do that, we follow a k-nearest neighbours rule [131] to derive the global position. Specifically, it ranks the 15 temporal positions of a specific feature in order. Among the ordered values, the first one that has k neighbours within an λ ms interval was picked as the final decision. Eventually, finalised temporal position of each feature from both manual annotations and

[9]The list of records selected from QTDB is: [sel30, sel31, sel32, sel33, sel34, sel38, sel39, sel40, sel41, sel42, sel43, sel48, sel49, sel51, sel52, sel17152, sele0612, sele0303, sele0409, sele0609, sele0612, sele0166, sele0211, sel223, sel301, sel16539, sel16786].

feature detection algorithms are obtained. On this basis, mean and standard deviation between the two can be derived for the 30 subjects.

3.4.1.2 Results and Discussion

Table 3.4 shows the results in terms of the mean μ and standard deviation σ for QTDB. With these two metrics, it is possible for us to have the standard to compare against the results. Also, in the first row, the *two-standard-deviation* tolerance defined by the CSE Working Party from measurements made by different experts are listed [138, 204]. In addition, to make it practically more understandable, equivalent number of samples of standard deviation is also listed for our algorithms. Due to the fact that QTDB applies 250 Hz sampling frequency, 1 sample of data corresponds to 4 ms. Therefore, samples of σ of both algorithm is obtained as being rounded to the closet integer. With this metric, we can have the sense of how many actual digitised samples are away from the CSE tolerance. Additionally, results from Martínez [138], Rincón [205], Jane [206] and Almeida [197] as competitor methods are also given in the table, in order to provide a comprehensive comparison with the proposed algorithms. Note that, the best results among these four methods for one feature are specifically boldfaced.

Now, the results of both algorithms will be compared with the best results obtained from the competitor methods. Note that, because one data sample in QTDB represents 4 ms. So as long as the difference between the best results and the proposed ones is within this threshold range, it can be regarded as "comparable". As shown in Table 3.4, we can see that only HFDA in P_{on}, HFDA in P_{peak}, TDMG and HFDA in QRS_{on}, TDMG in T_{on} exhibit comparable results to the best algorithms'. For the rest of the fiducial points, they show relatively poor $\mu \pm \sigma$. Note that R_{peak} can not be compared because no available algorithm generate this feature.

Apart from the comparison with exiting competitor methods, we also need to compare our results to CSE tolerance limits. To do that, we convert our results into the form of data samples, to calculate the difference of σ relative to the CSE limits. Doing so allows us to compare reasonably with the CSE tolerance. Under this setting it can be observed that, for TDMG the sample difference equals 0 in QRS_{on} and QRS_{off}, but with 2 in P_{off} and 3 in P_{on} and T_{peak}. For HFDA, the sample difference equals 0 in QRS_{on}, but with 1 in P_{on}, P_{off}, QRS_{off} and 4 in T_{peak}. With this amount of difference in samples, the algorithms work actually fine due to the fact that in real clinical practice, a few samples difference on display for clinical decision may not be much of difference.

Table 3.5 lists the results for PTBDB. As no available studies of feature extraction have been reported on this database, here we only focus on the features for which CSE tolerance limits are available. As a database, PTBDB allows for a more valid comparison with the CSE tolerance limits as it also contains standard 15-lead ECG signals. With

the process of k-nearest neighbours on extracting both global manual annotations and feature detection results, the best performance are obtained with k = 3 neighbours for all features, and $\lambda = 10$ ms for the P_{on}, P_{off} and QRS_{on}, while $\lambda = 12$ ms for QRS_{off} and T_{off}. According to the results in the table, both algorithms performs fairly similar as for mean of error. Regarding standard deviation, all CSE tolerance limits are satisfied apart from P_{on}, P_{off} for the first algorithm and QRS_{on} for the second.

3.5 Concluding Remarks

In this chapter, we have proposed two ECG fiducial points detection algorithms: one is based on time-domain gradient and morphology of the ECG with an aim of achieving accurate detection; the other is based on time- and frequency-domain hybrid approach with an aim of achieving low computational complexity while maintaining acceptable accuracy. In general, these methods both follow two stages when extracting specific features - both first operates initial estimation of the features, and then refinement of the initial results is executed by extra processing steps to improve the accuracy. In terms of detection performance, in general both algorithms achieve relatively poor detection accuracy in most of the features with respect to the state-of-the-art algorithms except a few features. In comparison with CSE limits, the algorithms works relatively good with the metric of data samples as in difference of σ.

Despite the poor performance of the proposed algorithms, a new feature named spectral energy will be given in later chapters. As we will prove via statical analysis and classification performance analysis, most of the spectral energy features exhibit strong robustness against misdetection generated from feature detection algorithms. That means, spectral energy will compensate the poor performance of the ECG fiducial point detection algorithms. In addition, as we will show in later chapters also, features namely P_{on}, P_{off}, QRS_{on}, QRS_{off}, T_{on}, T_{off} are of paramount importance for ECG normal and abnormal heartbeat classification. On this basis, we choose to take the first algorithm as the primary feature detection method for later analysis. This decision stems from the fact that it performs relatively better in QRS boundaries and T boundaries, though not in P boundaries.

Table 3.4: Fiducial Points Detection Performance of TDMG and HFDA on QTDB.

Method	**Feature**	P_{on}	P_{peak}	P_{off}	QRS_{on}	R_{peak}	QRS_{off}	T_{on}	T_{peak}	T_{off}
CSE [204]	2σ(ms)	10.2	/	12.7	6.5	/	11.6	/	/	30.6
Martínez	# ann. beats	3194	3194	3194	3623	/	3623	/	3542	3542
et al. [138]	$\mu \pm \sigma$(ms)	**2.0±14.8**	3.6±13.2	1.9±12.8	4.6±7.7	/	**0.8±8.7**	/	**0.2±13.9**	-1.6±18.1
Rincón	# ann. beats	3194	3194	3194	3623	/	3623	/	3542	3542
et al. [205]	$\mu \pm \sigma$(ms)	8.6±11.2	10.1±8.9	**0.9±10.1**	**3.4±7.0**	/	3.5±8.3	/	3.7±13.0	**-2.4±16.9**
Jane	# ann. beats	3194	3194	3194	3623	/	3623	/	3542	3542
et al. [206]	$\mu \pm \sigma$(ms)	14.0±13.3	**4.8±10.6**	-0.1±12.3	-3.6±8.6	/	-1.1±8.3	/	-7.2±14.3	13.5±27.0
Almeida	# ann. beats	/	/	/	3412	/	3412	1302	/	3331
et al. [197]	$\mu \pm \sigma$(ms)	/	/	/	7.5±11.2	/	6.1±12.3	**18.7±27.6**	/	7.9±21.7
TDMG	# ann. beats	1620	1620	1620	1620	1620	1620	1620	1620	1620
	$\mu \pm \sigma$(ms)	0.33±21.1	7.6±15	11.2±20.8	4.1±8.7	-5.2±15.6	5.1±12.4	12.1±24.6	2.8±25.3	5.6±28.6
	σ(samples)	5	4	5	2	4	3	6	6	7
	Diff of σ	3	/	2	0	/	0	/	3	/
HFDA	# ann. beats	1620	1620	1620	1620	1620	1620	1620	1620	1620
	$\mu \pm \sigma$(ms)	-6.3±12.5	5±9.5	3.1±16	3.7±7.8	3.8±9.8	12.1±16.6	-15.8±34	-15.3±29.3	-16.6±20.8
	σ(samples)	3	2	4	2	2	4	8	7	5
	Diff of σ	1	/	1	0	/	1	/	4	/

Table 3.5: Fiducial Points Detection Performance of TDMG and HFDA on PTBDB.

Method	**Feature**	P_{on}	P_{off}	QRS_{on}	QRS_{off}	T_{off}
CSE [204]	2σ(ms)	10.2	12.7	6.5	11.6	30.6
TDMG	# ann. beats	422	422	450	450	432
	$\mu \pm \sigma$(ms)	0.2±18.7	-1.5±22.4	-4.1±6.4	1.3±8.7	-11.9±19.8
HFDA	# ann. beats	422	422	450	450	432
	$\mu \pm \sigma$(ms)	1.1±9.5	-6±11	3.8±10.8	3.7±6.8	-8±10.8

Chapter 4

Spectral Energy as a Feature for Normal and Abnormal ECG Classification

Chapter 3 presented two ECG fiducial points detection algorithms based on time-domain gradient and morphology analysis of the ECG as well as time- and frequency-domain hybrid approach. They have been developed and finally shown to generate acceptable outputs. These outputs are considered to be clinically useful features. However, how to deploy the features in classification remains a question. Furthermore, misdetection error of these clinical features is inevitable. Simply deploying them in classification may not lead to satisfied performance due to error propagation along the way. With this in mind, we attempt to explore alternatives to conventional clinical features, so as to form a feature that is robust against the error encountered in fiducial points detection algorithm while being implementable at a wireless sensor node. In this chapter, we will explore spectral energy as such a feature.

The rest of the chapter is organised as follows. Section 4.1 presents the motivation behind the works in this chapter and Section 4.2 introduces the background of spectral energy and preliminaries of the classification algorithms. Section 4.3 outlines an investigation into four different methods of obtaining spectral energy from an entire ECG complex. This would explain how we are going to deploy the clinical features generated from the fiducial points detection algorithm. From there, we establish spectral energy of which ECG wave components is preferred. In Section 4.4, we then continue to explore the robustness of spectral energy against misdetection error of the wave boundary from two perspectives – one is statistical analysis of the variation of spectral energy under misdetection, the other is classification performance using spectral energy as a feature under misdetection. Finally, Section 4.5 concludes the chapter.

4.1 Motivation

In mobile application, the complexity of signal processing needs to be low to preserve battery life. As our main target is to classify normal and abnormal ECG in mobile environment, preferably the number of features involved in this process should be small while at the same time producing acceptable outcome of classification. In principle, a certain physiological condition is expected to make certain morphological changes in different components of the ECG wave. Therefore, in general, features based on morphology of the waves are used to analyse and make further prediction on physiological conditions. As a result, sophisticated and hence computationally intensive signal processing algorithms are used to capture and extract these changes in time domain [1]. To achieve high performance, the number of features required may be large. That in turn implies that classification with these features may be computationally intensive.

In addition, such morphological changes may not always be evident in pure visual inspection. However, if morphology changes, technically the change will also affect the frequency-domain characteristic. This in fact prompts us to think that spectral feature may be used for detecting such morphological changes. This also means that it has potential for a new feature that could be applicable for our purpose, as it may so happen that frequency domain based processing exhibit less computational complexity, while being able to reflect the morphological changes. As a result, analysing the ECG signal in frequency domain may provide a computationally less expensive alternative to ECG classification.

Apart from that, finding a way to tackle the misdetection error generated from feature detection algorithm is also important. This is because such error may propagate down to classification, inflicting inferior performance upon classification. It may be even worse when we deploy simple, low-complex classification algorithms like LDA and QDA, as they do not have the capability of rectifying subject lying on *wrong* side of the decision boundary like SVM does.

To resolve our concerns, spectral energy as a feature is considered in our work. Especially, spectral energy of different ECG wave components is considered. To obtain spectral energy, first we need to identify the onset and offset points of different ECG waves. However, automated algorithms have their own limitation in calculating the onset and offset points, as we have already discussed in Chapter 3, and thereby misdetection error may be introduced to these points. This is also true when standard clinical features are used, namely QRS interval, R-R interval, etc. As a result, we first explore the applicability of spectral energy as a feature for classifying normal and abnormal ECG. Then, we explore its robustness (or consistency) against wave boundary misdetection to take into account the possible limitation of the automated ECG fiducial point detection algorithm.

4.2 Background

4.2.1 Definition of Spectral Energy

As we have already discussed in Section 4.1, spectral energy has been hypothesised to act as a feature for ECG classification. But before we proceed to any further experimental investigation, definition of spectral energy should be clarified and will be given in the following.

In digital signal processing, deterministic signals are signals whose nature are completely specified for any given time. It can therefore be modelled by a specific function of time [207]. Given a deterministic signal g(t), its Continuous-Time Fourier Transform (CTFT) can be defined by

$$G(f) = \int_{t=-\infty}^{\infty} g(t)e^{-j2\pi ft}dt \tag{4.1}$$

The total energy of the signal can be given by

$$E = \int_{t=-\infty}^{\infty} |g(t)|^2 \, dt \tag{4.2}$$

According to Parsevals theorem, the total energy can also be obtained from the amplitude spectrum of the signal, which is

$$E = \int_{t=-\infty}^{\infty} |g(t)|^2 \, dt = \int_{f=-\infty}^{\infty} |G(f)|^2 \, df \tag{4.3}$$

Here, it is possible to observe that G is devoted to the contribution of the total energy of the signal in a frequency band $(f, f+\Delta f)$. Integrating over the entire frequency band leads to the total energy. As a result, $|G(f)|^2$ itself is regarded as the Energy Spectrum Density (ESD) of the signal g(t).

In addition, as for a discrete time series $x[n]$, we observe $x[n]$ from 0 to N and assume that the samples of $x[n]$ outside this interval equal to zero. Since the signal is truncated and zero padded outside the interval, its energy is finite and thus it satisfies the condition of Discrete Fourier Transform (DFT) and is therefore applicable to DFT. As a result, DFT of the observed $x[n]$ can be obtained as

$$X(f) = \sum_{n=0}^{N} x[n]e^{-j2\pi fn/N} \tag{4.4}$$

In accordance with Equation 4.3, again $|X(f)|^2 \Delta f$ in this case can be viewed as the energy of the truncated signal that is contributed by the frequency components between

f and $f + \Delta f$. It has, therefore, given a way of calculating the spectral energy. By using DFT, ESD of the signal throughout the whole frequency span can be obtained. The spectral energy within a frequency range of our interest can be computed as

$$E_F = \sum_{f=f_1}^{f_2} \left| X(f)^2 \right| \tag{4.5}$$

Equation 4.5 states a way of computing the spectral energy using DFT. So on the other hand, what about DWT? As we know, in DWT the mother wavelet function is discretised (Equation 2.10), and therefore constitutes the orthonormal basis. That means, the idea of energy can be linked with the one derived from the Fourier theory [208]. Moreover, by using DWT it is possible to not only obtain the spectral energy within a particular frequency range, but also localise the temporal span of spectral energy due to its nature for non-stationary signal analysis. As we discussed in Chapter 2, DWT produces approximate coefficient and detail coefficient. To retrieve the spectral energy of frequencies at a certain decomposition level (level j), essentially detail coefficient is utilised. From that, calculation of spectral energy based on DWT can be given in Equation 4.6, in accordance with Equation 2.11.

$$E_W^j = \sum_{n=n_1}^{n_2} |cD_j[n]|^2 \tag{4.6}$$

4.2.2 Related Works

As discussed in Section 4.2.1, spectral energy contained in a signal within a particular time frame can be obtained by summing across all of its frequency components in squared term. Same approach can also be applied to achieve energy of a specific frequency range.

To compute the spectral energy of an ECG signal, there are two main approaches: DWT and DFT. Successful examples of using spectral energy have been reported in a few works. On one hand, spectral energy of ECG has been used to classify normal and Myocardial Infarction (MI) ECG signal [140]. It has been stated that spectral energy was computed from DWT detail coefficient vectors at level 2 to 4 and approximate coefficient vector at level 4 respectively with Daubechies 6 (or, db3[1]) mother wavelet, and then considered as features for classification. Moreover, [209] has also stated that, by using Haar, db6 and bi-orthogonal wavelet as the three mother wavelets, mean and standard deviation of the spectral distribution of energy between normal and MI patients were extracted and acted as the inputs to neural network. It claimed that, although these two features are simple, they effectively classify normal and abnormal cases with more than 75% accuracy based on Receiver Operator Curve (ROC) analysis. On the other

[1] Daubechies 6 refers to the length, while db3 refers to the number of vanishing moments.

hand, Fourier Transform based spectral energy has been successfully deployed in [210] for ECG signal classification. Non-parametric power spectrum estimation methods, namely periodogram, modified periodogram, Welch method and multitaper method, were utilised to produce 129 spectral features, respectively. In addition, three timing interval feature were also considered as features as well. All were combined as feature vector and thereafter classification of five types of ECG beats was carried out with SVM classifier using Gaussian radial basis function (RBF).

4.2.3 Methods for Calculating Spectral Energy

Although spectral energy has been shown to be a distinctive feature in ECG classification, deferent ways that spectral energy may be computed from an ECG give variety of possibilities in effecting the final performance of ECG classification. Instead of considering the spectral energy of the entire ECG as a feature, as all of the works discussed above did, we propose to use spectral energy of each wave component of the ECG as separate features. This intuitively results from the point of view that, different classes of ECG anomalies that are directly related to cardiac event are reflected upon different waves - P, QRS and T. Therefore, individual characteristics of each of them may have more discriminative property for classifying normal and abnormal ECG compared to the spectral energy of an entire ECG complex. Additionally, thanks to the very nature of WT, the time and frequency information of each wave component can be retained. It therefore allows us to observe both at the same time and retrieve spectral energy for our purpose.

Nonetheless, to acquire the spectral energy of each wave component, wave boundary detection algorithm is needed. This in fact reminds us of an issue: from computational complexity point of view, execution of this type of algorithm in mobile environment would consume comparative portion of power, hence reducing the battery life. Paradox may then occur if spectral energy of wave components derived as a set of features for classification could not reach our desired performance. In this case, it was not worthwhile to acquire this set of features at the expense of running the wave boundary detection technique as the preceding step. But rather it might be possible to see that, the spectral energy of the entire PQRST complex might be proved to be outstanding enough to achieve comparative level of accuracy as the combined effort of wave boundary detection and spectral energy of individual wave components does, or even better. In such case, wave boundary detection algorithm would not be necessary and therefore could be eliminated. Thereby, workload could be reduced and further save the battery life.

To make an exploratory comparison, we need to find some ways to calculate the spectral energy of the entire PQRST complex. By then we can compare them with our proposed approach. Here, DFT and DWT will be used to simply calculate the spectral energy of

the entire PQRST complex. But before we start the experiments, a brief comparison between DFT and DWT is given below. Note that, different mother wavelets may lead to varied number of arithmetic operations in DWT analysis, thus the corresponding computational complexity. So, to give an idea of the difference of computation complexity between the two, time complexity is used here.

- *DFT*: To compute DFT in practice, Fast Fourier Transform (FFT) is normally used to implement computationally efficient design. Assuming that the size of the signal is n, the time complexity of running FFT takes $O(n \log n)$ [211].
- *DWT*: Trade-off of computational complexity can be made, as less complex but appropriate mother wavelet (e.g. Haar wavelet) can be selected to perform DWT. Again, assuming that the size of the signal is n, the time complexity of running DWT with dyadic grid of resolution reaches $O(n)$ [212].

As we can see from time complexity, with a bigger n the complexity of FFT is nonlinearly becoming bigger than DWT. Therefore, FFT generally requires more time and thus more arithmetic operations than DWT to finish.

Apart from these two approaches, another method of retrieving the major frequency components, hence the majority of the energy of PQRST complex, may be possible as well. Here, DWT-based thresholding policy is introduced and applied, with hope of exploring a simple, low-complex and efficient method to compare with our proposed approach. This method will be discussed in detail in Section 4.3.

Overall, three different ways of acquisition of spectral energy are considered to challenge our proposed method. Eventually this leads us to four sets of experiments. They have been set up to explore the possible acquisition of spectral energy as a feature, and the performance of classification coupled with classifiers is considered as the primary evaluation metric.

4.3 Spectral Energy as a Feature

In this section, four different approaches of deriving spectral energy of an entire ECG complex are covered here. A brief experimental strategy we take is explained below. Note that, all the following experimental strategies are coupled with our selected classification algorithms to achieve classification. The more detailed principle of each of these strategies is described in the following sections. **Experimental strategy**

- *Set 1 – DFT based spectral energy of the entire ECG complex*: spectral energy of the entire PQRST complex is derived by the conventional signal processing technique – DFT.

- *Set 2 – DWT based spectral energy of the entire ECG complex*: spectral energy of the entire PQRST complex is derived by DWT.
- *Set 3 – DWT and thresholding based spectral energy of the entire ECG complex*: a DWT and thresholding policy based coefficient selection approach is applied to obtain key spectral coefficients of the signal, which then are used for spectral energy derivation.
- *Set 4 – DWT based spectral energy of the ECG wave components*: Each of the key ECG wave components is extracted and analysed by utilising the time-frequency domain feature of DWT. Thus, according to their specific frequency characteristics, spectral energy of each wave component is derived.

4.3.1 Database

For the experimental analysis we utilised ECG excerpts from two databases. In total we selected 104 12-lead records divided in two categories (normal/abnormal). For each of these records full clinical diagnosis existed. Specially, we used PTB database (PTBDB) available in PhysioNet [201]. It contains standard 15-lead ECG recordings at 1 KHz covering various disease categories. The 52 records for the normal class (healthy control) were obtained from this database. From PTBDB we also selected 17 records diagnosed with myocardial infarction covering all available (anterior, inferior, lateral and posterior) subclasses, while 18 records were equally collected from the other six disease classes, namely cardiomyopathy, bundle branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease and myocarditis. To equalise the number of normal and abnormal records, 17 ECG signals from patients with diagnosed myocardial scar were obtained from the Southampton General Hospital Cardiology Department (SGHCD) database. These records are standard 12-lead paper ECG, sampled at 500 Hz which were digitised at a rate of 1 KHz with the use of the ECGScan software [213]. The reason for selecting scar patients is that, patients who have scar are more likely to evolve into danger if he/she somehow has heart attack again after the rebuilding of infarcted myocardium [214]. Therefore, we also consider patients with scar that is clinically examined and reported in SGHCD. Finally, since the 17 records from SGHCD had only 12-leads, we considered the same 12-leads for the PTBDB records.

Note that there are 12 lead ECG signal available for each record. From each of these leads we extract one full PQRST complex, with same time frame as defined in Section 3.2.3.2. But since we are only interested in single heartbeat, excerpt was done by deploying heartbeat segmentation technique (see Appendix A), with assistance of manual adjustment to assure that complete heartbeat is covered. Also note that, in a conventional clinical setting, 12-lead ECG is normally used where each lead provides one ECG trace. Clinical decision is made by observing all of them in parallel. In general, at

maximum 5 leads are used in such a setting [215]. Therefore in this work we restricted ourselves in exploring classification accuracy with combination of feature sets for up to 5 leads only. In particular, lead scenario refers to the number of participating leads under consideration.

4.3.2 Set 1: DFT based Spectral Energy of the Entire ECG Complex

Set 1 follows Equation 4.5 to acquire the spectral energy of the entire PQRST complex. It is worth pointing out that, we have constrained the frequency range to be 0.5 to 40 Hz. This is because of the fact that frequency components of the ECG waves (P, QRS, T) largely fall into this spectral region, as shown in Figure 4.1[2]. It is reasonable to only extract that particular part of energy of the signal, which in turn eliminates any interference from outside the region, e.g. powerline noise, high frequency noise, etc.

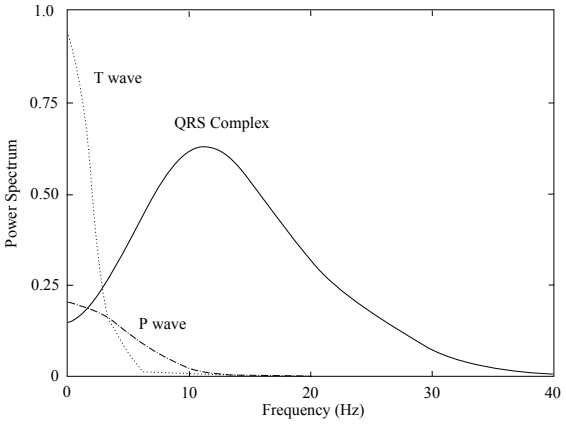


Figure 4.1: Power spectrum of the P wave, QRS complex and T wave.

With spectral energy computed, exhaustive simulation is conducted with the classifiers. Specifically, since there are only 12 leads and from each lead there is only one spectral energy of the entire PQRST complex as a feature, there are in total 12 features for each record in our database. So, exhaustive attempts of all combinations of leads up to 5 out of 12 leads (i.e. C_5^{12}) can be easily achieved. By doing so, the highest classification accuracy can be observed using 10-fold CV in each lead scenario (1 to 5 lead scenario) for the five classifiers. Table 4.1 shows the best raw classification accuracy (Acc) in each lead scenario. Note that the highest performance in certain lead scenario among the five classifiers is highlighted in bold. In addition, arrangement of participating lead in each lead scenario can be found in Appendix B. The rest of the experiments follow the same convention as Set 1 in terms of results and lead arrangement.

[2]This figure is reproduced according to Figure 6.11 in [49]. Note that small variations between heartbeats may exist.

Table 4.1: The best raw classification accuracy (Acc) in each lead scenario for the five classifiers for Set 1 experiment.

Lead	**Set 1**				
Scen.	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
1	69.23	68.27	69.23	70.19	**71.15**
2	75.96	75.00	**76.92**	**76.92**	75.00
3	80.77	77.88	**81.73**	78.85	77.88
4	80.77	78.85	81.73	**82.69**	79.81
5	79.81	79.81	81.73	**82.69**	79.81

4.3.3 Set 2: DWT based Spectral Energy of the Entire ECG Complex

To keep consistency of the entire work flow as we did in Chapter 3, Haar is again selected as the mother wavelet. More technical details and demonstration of Haar DWT analysis can be referred back to Section 3.3.3.1. In addition, the experimental set-up for Set 2 may be divided into two cases where two ways of obtaining spectral energy of the entire PQRST complex are considered.

- Case 1 with spectral energy based on decomposition level 1 to 5
- Case 2 with spectral energy based on decomposition level 3 to 5

Since level 1 and 2 mainly contains the high frequency components and hence the possibility of existence of noise, it has been eliminated in Case 2 just to examine the impact of the spectral energy derived from these two levels. Both cases follow Equation 4.6.

Exhaustive simulation is conducted with the classifiers fed by the spectral energy computed from both cases, respectively. Again in total 12 features for each record are obtained. Through the exhaustive simulation considering up to 5 leads out of 12, the highest classification accuracy in each lead scenario can be achieved using 10-fold CV for the five classifiers. The associated results can be found in Table 4.2.

4.3.4 Set 3: DWT and Thresholding based Spectral Energy of the Entire ECG Complex

In essence, spectral energy calculation is all about selection of proper frequency coefficient. By acquiring useful spectral coefficients, the majority of signal energy can be obtained. A simple and effective approach of attaining the significant coefficients while disregarding the remaining redundant ones would be very much preferred. Therefore, the less computational complexity required by the system is, the longer battery life can be.

Table 4.2: The best raw classification accuracy (Acc) in each lead scenario for the five classifiers for Set 2 experiment.

Lead	**Set 2 Case 1**					**Set 2 Case 2**				
Scen.	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
1	67.31	70.19	67.31	**73.08**	67.31	69.23	**75.00**	69.23	74.04	71.15
2	76.92	76.92	78.85	**80.77**	76.92	77.88	76.92	76.92	**79.81**	76.92
3	79.81	75.96	78.85	80.77	**81.73**	79.81	75.96	78.85	**80.77**	80.77
4	80.77	81.73	79.81	80.77	**85.58**	80.77	81.73	79.81	80.77	**86.54**
5	81.73	76.92	80.77	81.73	**85.58**	81.73	76.92	80.77	81.73	**84.62**

In the domain of ECG compression, great efforts have been done in investigating the trade-off between the degree of compression and the level of fidelity of the reconstructed ECG signal [216, 217]. Same philosophy may also apply to our spectral energy calculation, as the fundamental idea behind is to investigate an effective way to minimise the number of samples. In [217], a DWT-based low-complexity ECG compression method is proposed. From there, only those coefficients which have significant contribution to the total energy of the original signal is chosen by a thresholding policy, thereby achieving a good trade-off. Inspired by this fact, a DWT and Thresholding based coefficient selection approach is applied here. Different to [217] though, instead of deploying energy packing efficiency, we opt to use classifier performance to judge the performance of the threshold selection in our work. In the following, detailed description of our approach is given.

Specifically, at first the entire PQRST complex is subjected to DWT analysis. The algorithm then captures the maximum (i.e. positive maximum) and minimum (i.e. negative maximum) values of the coefficients within the entire time window of the signal at a specific level, particularly level 1 to 5. By the time we have picked up these two most dominant frequency components in each level, threshold is applied. That means, coefficient that has a higher or lower magnitude from a certain percentage (i.e. threshold) of the maximum or the minimum value, respectively, are considered to be significant and therefore retained; for the rest of coefficients, they are redundant and simply ignored. Note that such percentage refers to the percentage level, defined by threshold, of maximum or minimum value. Once significant coefficients are obtained, the spectral energy can then be derived.

With this in mind, exhaustive simulation is done in a way that variety of thresholds, ranging 1%, 2%, 4%, 8%, 16%, 32% is used in different levels to filter out the coefficients for final spectral energy calculation. To put it in a systematic way, we start from tuning the threshold of level 1 ($Thres_{lv1}$) from 1% to 32%, while keeping the threshold of other levels ($Thres_{other}$) to be 1%. 10-fold CV is deployed and the corresponding classification results are obtained. After that, threshold of level 2 ($Thres_{lv2}$) is tuned from 1% to 32%

Table 4.3: Threshold setting: tuned level (TL), threshold percentage (TP) and other threshold percentage (OP) in each lead scenario for Set 3 experiment.

Lead	**LDA**			**QDA**			**SVM$_L$**			**SVM$_Q$**			**k-NN**		
Scen.	**TL**	**TP**	**OP**	**TL**	**TP**	**OP**	**TL**	**TP**	**OP**	**TL**	**TP**	**OP**	**TL**	**TP**	**OP**
1	4	32%	32%	5	4%	32%	1	32%	32%	4	8%	32%	4	4%	32%
2	3	1%	2%	3	4%	4%	4	1%	16%	2	32%	16%	4	1%	16%
3	5	16%	4%	2	1%	4%	3	4%	4%	2	4%	8%	5	8%	2%
4	3	2%	1%	3	8%	1%	1	16%	1%	1	8%	16%	3	1%	32%
5	2	1%	8%	5	32%	2%	2	4%	1%	4	32%	32%	2	16%	4%

Table 4.4: The best raw classification accuracy (Acc) in each lead scenario for the five classifiers for Set 3 experiment.

Lead	**Set 3**				
Scen.	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
1	68.75	74.52	69.04	**75.38**	69.81
2	77.60	76.44	77.40	80.00	**80.19**
3	80.38	78.27	79.13	82.21	**82.40**
4	80.58	80.00	79.42	81.15	**84.23**
5	80.29	81.25	79.90	83.17	**87.31**

instead while $Thres_{other}$ to be 1% as before. Classification results are again obtained in this instance. The entire process is repeated for $Thres_{lv3}$, $Thres_{lv4}$ till $Thres_{lv5}$ is tuned. Then the process starts again from the beginning, but with $Thres_{other}$ changed to be 2%. This is repeatedly done until $Thres_{other}$ is ultimately set to 32%. Overall, by doing so, the major possible selection of thresholds for each level is covered. With proper thresholds, the dominant frequency components are expected to be captured and the associated classification results of each classifier can be achieved.

Furthermore, this approach is applied in every possible lead combination and lead scenario to explore the accuracy derived from this set of experiment. As we only concern about the highest accuracy that possibly produced in each lead scenario for the 5 classifiers, others are simply ignored. Having done that, it can be known that what combination of threshold in tuned level and in other levels would produce the best accuracy provided by a specific classifier. Table 4.3 shows the threshold setting as in the optimal tuned level (TL) that is chosen to tune, the optimal threshold percentage (TP) for this level, and the optimal threshold percentage for other levels (OP). Following is Table 4.4 where the best raw classification accuracy with the associated combination of thresholds from Table 4.3 in each lead scenario is shown. Note that k-NN in all lead scenarios except 1 lead scenario shows superior performance than the rest, particularly SVMs.

Though technically SVMs are perceived to surpass k-NN in terms of classification performance and robustness, k-NN may actually surpass SVMs when data distribution is of low sparsity [218].

4.3.5 Set 4: DWT based Spectral Energy of the ECG Wave Components

Rather than acquiring the spectral energy of the entire PQRST complex, Set 4 is designated to take advantage of DWT in order to get the spectral energy of certain wave components of interest. More detailed explanations are given in the following subsections.

4.3.5.1 Wave Component Detection

Once a 12-lead ECG signal is available, the boundaries of the wave components of our interest can then be extracted using our automated feature detection tool discussed in Chapter 3. As a result, the onset and offset of P/QRST/T waves can be obtained directly and utilised as input for the next stage.

4.3.5.2 Spectral Energy Computation for Wave Components

Once isolated, DWT with Haar as the basis function is applied to the entire PQRST complex in order to obtain the detail coefficients for later spectral energy calculation. Decomposition level 2 and 3 of DWT effectively show the high-frequency components (e.g. QRS complex) of the ECG signal, while level 5 mainly corresponds to low-frequency components (e.g. P and T waves). Note that, although previously in Section 3.3.3.1 and 4.3.3 we stated that level 2 mainly exhibits noisy signal, this time we include level 2 into our analysis because (1) from feature detection perspective, level 2 is noisy and not suitable in that it barely provides distinct shapes of the coefficient sequence between different waves for analysis in order to generate accurate results; but (2) from spectral energy perspective, the way that measures spectral energy makes noise less effective in level 2, not to mention higher levels.

Moreover, the time-frequency localisation property of DWT is utilised here for isolating the frequency components corresponding to the P wave, QRS complex and T wave. So, by applying Equation 4.6 on the DWT coefficients from level 2 and 3 respectively, the QRS complex feature is obtained leading to two spectral energy features QRS_2 and QRS_3. On the other hand, the coefficients at level 5 is used to generate three more spectral energy features – P_5, T_5, PR_5 for P wave, T wave and PR interval, respectively. Regarding the QT-interval, which contains both high and low frequency components,

we calculate its spectral energy by summing the individual QT-interval spectral energies from coefficients in level 3 and 5, as well as coefficients in level 3, 4 and 5 to produce two different calculations of the QT-interval spectral energy – QT_{345} and QT_{35}. The reason for including level 4 in QT_{345} is to capture the spectral energy corresponding to the transition from high to low frequency. So, in total 7 distinct spectral energy-based features were calculated on a per lead basis and grouped into three categories:

- *low-frequency feature group (L)*: P_5, T_5, PR_5;
- *high frequency feature group (H)*: QRS_2, QRS_3;
- *combined high and low-frequency feature group (B)*: QT_{345}, QT_{35}.

4.3.5.3 Feature Selection and Acquisition of Best Feature Space

We start our exploration using all the 12-lead ECGs available to us and extract one full PQRST complex from each of the leads. Each of these isolated PQRST complex undergoes the feature generation procedure described in Section 4.3.5.2, resulting in 7 distinct spectral energy features per ECG beat and totalling to $7 \times 12 = 84$ features. Our aim is to find out the best set of features through simulation using different lead combinations in this feature space under the constraint of limited number of available leads, which may give the best classification result. To do that, feature ranking and feature space selection as the two key steps need to be invoked.

Feature Selection One way to ascertain the best feature combinations from a pool of potentially useful features is by means of exhaustive simulation. However, due to the fact that exhaustive simulation in this large feature space is extremely time intensive, feature selection should take place beforehand so that the most relevant features may be obtained. Fisher's criterion [167] is thus deployed in our work to select one feature from the L, H and B frequency groups for each lead, which can separate the two classes (i.e. normal and abnormal ECG) for each lead to the maximal extent. By doing so, it is expected to achieve the most discriminating features for classification. Equation of Fisher's criterion can be given as

$$J = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{4.7}$$

where m and σ^2 denotes the mean and variance of each class [166]. In essence, Fisher's criterion calculates the ratio of the between-class variance to the within-class variance on the basis of one feature and indicates the extent of mean separation and overlap between the two classes. Therefore, once the ratios for 84 features are obtained, the

Table 4.5: The most discriminant energy feature in each frequency group for each lead.

	Lead (Abbr)	**I (1)**	**II (2)**	**III (3)**	**aVR (4)**	**aVL (5)**	**aVF (6)**
Feature Group	**L**	T_5	T_5	T_5	T_5	T_5	T_5
	H	QRS_2	QRS_3	QRS_2	QRS_3	QRS_3	QRS_2
	B	QT_{345}	QT_{345}	QT_{35}	QT_{345}	QT_{35}	QT_{345}
	Lead (Abbr)	**V1 (7)**	**V2 (8)**	**V3 (9)**	**V4 (10)**	**V5 (11)**	**V6 (12)**
Feature Group	**L**	P_5	P_5	PR_5	T_5	T_5	T_5
	H	QRS_3	QRS_2	QRS_2	QRS_3	QRS_3	QRS_2
	B	QT_{345}	QT_{35}	QT_{35}	QT_{35}	QT_{35}	QT_{345}

most distinctive feature for each feature frequency group can be determined by selecting the highest one within that frequency group for each lead. The final selected features for each lead under this principle are shown in Table 4.5. Note that the results in this table come from all the PQRST complexes for all patients.

Acquisition of Best Feature Space After deriving the best features, we used exhaustive simulation technique to find out the best combination of the leads out of the available 12 leads that may enable us to attain the maximum classification accuracy. Considering l number of leads out of the total of 12 leads, the number of possible lead combination is C_l^{12}. In each of these lead combinations, we opt to select at least one (at most three) feature from each individual lead to build up the feature space for performance assessment. In addition, as mentioned earlier, we restrict ourselves to the combination of 5 leads only to be consistent with the constraints imposed by the application scenario. All the samples of each feature are normalised (Equation 2.12) with respect to their mean and standard deviation at the beginning of our exploration and 10-fold CV is employed in this study. The five classifiers discussed in Section 2.3.4 are deployed here as the main classification models for ECG classification to incorporate with wrapper method. Training data has been used for optimising the parameters in the parametric models of the classifiers. Also, regarding SVM, the regularisation parameter C_s of SVM is set to 1. Conventional quadratic programming solving method is selected in the training phase for SVM. Notice that, for evaluating the performance of each of the classifiers we adopted the raw classification accuracy Acc.

By running exhaustive simulation and evaluating the results for each of the classifiers under consideration, we select the optimal lead combination and its associated feature combinations exhibiting maximal accuracy for every lead scenario. To do so, initially only one lead $l = 1$ is considered, and from all 12 leads the one that achieves maximum

Table 4.6: Feature space selection.

	Lead Scen.	**(Lead, Feature) Combination**
LDA	**1**	(3,LHB)
	2	(2,LH), (3,LB)
	3	(2,LHB), (3,LHB), (7,LB)
	4	(2,LHB), (3,LHB), (7,LB), (8,L)
	5	(1,L), (2,LB), (3,LHB), (7,B), (8,L)
QDA	**1**	(4,LB)
	2	(3,LHB), (4,LHB)
	3	(3,LH), (4,LB), (5,LHB)
	4	(3,LHB), (4,LB), (5,LHB), (10,H)
	5	(3,LH), (4,HB), (5,HB), (6,L), (10,LH)
SVM$_L$	**1**	(2,LH)
	2	(2,L), (3,LHB)
	3	(2,L), (3,LH), (8,LHB)
	4	(2,L), (3,LHB), (5,HB), (8,LB)
	5	(2,L), (3,LB), (5,B), (6,LHB), (8,LHB)
SVM$_Q$	**1**	(4,LH)
	2	(4,L), (8,LHB)
	3	(4,L), (5,LHB), (8,LHB)
	4	(3,LH), (4,L), (5,L), (8,LHB)
	5	(2,L), (3,LB), (4,L), (5,L), (8,LHB)
k-NN	**1**	(4,L)
	2	(4,L), (5,L)
	3	(4,LHB), (5,LHB), (9,H)
	4	(4,LHB), (5,LH), (8,B), (9,HB)
	5	(3,LHB), (4,LB), (5,LHB), (8,B), (9,H)

accuracy is selected for each classifier. Following, we keep this lead and couple it with each one of the remaining 11 in order to identify the best lead combination for $l = 2$. This process is repeated up to $l = 5$. By doing so, we assure that there is always improvement in the performance of a classifier as we add extra leads and thus their associated distinctive features.

Table 4.6 shows the lead combinations with which the classifiers could obtain the maximal performance in all the lead scenarios under consideration. The '(Lead, Feature) Combination' column shows which features from Table 4.5 in each feature frequency group for each lead are used. Clearly, each of the combination associates with its classification accuracy, as shown in Table 4.7, and it will be discussed in Section 4.3.6.

Table 4.7: The best raw classification accuracy (Acc) in each lead scenario for the five classifiers for Set 4 experiment.

Lead	**Set 4**				
Scen.	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
1	75.29	73.85	75.29	**81.76**	80.06
2	83.56	76.73	85.83	**88.75**	87.44
3	**88.94**	83.46	85.87	87.76	86.41
4	88.56	83.65	86.57	**89.74**	86.35
5	89.52	83.46	86.79	90.13	**90.58**

4.3.6 Comparison and Discussion

Having obtained raw classification accuracy (Acc) for each set of experiment, comparison and discussion for these four sets can be given as follows. Firstly, as for Set 1 (Table 4.1), in general the performance of each classifier increases as the number of leads increase until it reaches 4 lead scenario. Increasing the number of lead does not improve the performance too much, as each classifier saturates respectively at the fairly same level of accuracy beyond 3 lead scenario. The best performance we can achieve in this set is 82.69% by SVM$_Q$ in 4 and 5 lead scenario.

Secondly, it can be seen that Set 2 Case 1 and Case 2 (Table 4.2) exhibit fairly similar raw classification accuracy. Only one major difference can be observed in 1 lead scenario, where 2-5% gap exist between classifiers in each case. Beyond 1 lead scenario, not much difference can be observed for each of the classifier. This shows that including spectral energy from level 1 and 2 to the entire spectral energy of the signal do not improve the performance of any classifiers. Rather, it fairly deteriorates the performance in 1 lead scenario. Mainly it might be due to the fact that high frequency components are not the major frequencies in ECG signal and noise effect is therefore introduced to each data sample during validation process. The best performance we can achieve in this set is 86.64% by k-NN in 4 lead scenario. In addition, between Set 1 and Set 2, it can be observed that only maximally around 2% difference exists between respective classifiers from both sets in each lead scenario. This implies that, spectral energy derived from DWT actually exhibits similar distribution of data compared to DFT, hence not much difference in classification performance can be observed.

Thirdly, it can be seen that with the thresholding policy in Table 4.3, in general the performance of all classifiers in Set 3 (Table 4.4) increases as the number of leads increases. Interestingly, depending on the classifier and lead scenario, TL, TP and OP derived from exhaustive simulation vary from case to case. This also implies that, specific thresholding pattern should be deployed in order to achieve the highest performance given a specific classifier and lead scenario. Overall, the greatest improvement along with the increasing

number of leads can be observed in k-NN, and it reaches the best performance among all classifiers at 87.31% in 5 lead scenario.

Finally, regarding Set 4 (Table 4.7), it can be seen that no matter which lead scenario it is, all classifiers performs markedly better compared to Set 1, 2 and 3, with 6-9% in LDA, 2-7% in QDA, 5-7% in SVM_L, 7-12% in SVM_Q and 3-12% in k-NN. In particular, 5 lead scenario in Set 4 exhibits the highest performance throughout all classifiers. This can be explained by the fact that characteristic features based on the ECG wave components serve more effectively and distinctively than the overall spectral energy of the entire PQRST complex and thresholding policy.

In summary, from each of the set stated above, they all provide a unique way of deriving the spectral energy from an ECG heartbeat: either the spectral energy of the entire complex, spectral energy of wave components, or spectral energy of major frequency components. It can be seen that DWT-based spectral energy on specific ECG wave components outperforms either DWT-based and FFT-based on the entire PQRST complex, or approach of obtaining DWT coefficients using adaptive threshold method. On one hand, for all the five classifiers, the highest accuracy that can be achieved by Set 4 for each lead scenario are significantly better than Set 1 and 2. On the other hand, the optimal thresholds for fetching the dominant DWT coefficients in each lead scenario are acquired in Set 3, which in turn result in the spectral energy and hence the associated classification accuracy. Even then, Set 4 still performs better than Set 3 in all cases. Both of the arguments strongly lead us to believe that, with assistance of feature extraction algorithm, Set 4 is of high accuracy to be applied in our ECG classification. Thus, it strengthens the point that spectral energy of ECG wave components can be considered as an interesting and promising feature. Therefore Set 4 has now been chosen as our fundamental method for extracting the spectral energy from the ECG PQRST complex.

4.4 Robustness of Spectral Energy

In Section 4.3, we established the potential of wave components based spectral energy for classifying normal and abnormal ECG. But still, it is a process that requires ECG wave boundary detection and therefore misdetection error generated by the automated algorithm cannot be avoided (see Chapter 3). It may be even possible that such error could affect how spectral energy is constructed and its associated accuracy, and further jeopardise later phase of classification. To investigate how spectral energy responses to the misdetection and whether it is a feature suitable for normal and abnormal ECG classification under misdetection, we need to explore in the following ways.

Experimental strategy

- *Artificial error injection*: As a pre-processing step, it is a way of injecting certain amount of error to the temporal location of wave boundary in different situations. With this, all possible misdetection situations with certain error can be investigated.

- *Statistical analysis of the variation of spectral energy under misdetection*: With artificial error injection, statistical analysis will be carried out. Mean bias (or mean difference[3]) between spectral energy under worst-case misdetection error and spectral energy of ground truth (based on cardiologists annotation of the wave component), its 95% confidence interval and limits of agreement will be derived and discussed to examine the robustness of spectral energy with worst-case misdetection error.

- *Classification performance using spectral energy as a feature under misdetection*: With artificial error injection and our proposed modified 10-fold CV, classification performance using spectral energy will be investigated to examine the robustness of spectral energy under misdetection.

4.4.1 Artificial Error Injection

To derive the spectral energy of a specific wave component, the corresponding onset and offset (i.e. wave boundary) are crucial. However, as discussed in Section 4.1, issues may arise where errors are introduced during feature detection process, leading to problematic spectral energy and further jeopardising the classification performance. Therefore, it is worth investigating how error is generated from the ECG feature detection algorithm and further the associated effects reflected upon the classification performance.

To do that, artificial error injection is carried out, thereby misdetection error can be simulated. Figure 4.2 depicts a simple approach to inject artificial error to the boundary detection of an ECG wave, in this case QRS complex as an example. Similarly, same principle can be applied to other wave components of interest. Firstly, the correct onset and offset of QRS wave is given by the cardiologist. These two parameters are regarded as ground truth. The reason of using the annotations from cardiologist[4] is because it is of our interest to see, having misdetection taken into account, how spectral energy derived would deviate from the one based on the ground-truth wave boundary. Secondly, eight cases of possible boundary misdetection are considered, which exhibit eight different combinations of misdetection deviating to either one or the other direction. To state the degree of misdetection error, Temporal Index of Misdetection (TIM) in millisecond

[3] mean difference is more towards a statistical term.

[4] Because only one cardiologist from Southampton General Hospital was available to annotate the ECG heartbeats, no inter- or intra-annotator variability analysis is taken into account in this study. However, such kind of analysis and solution has to be made when more than one human annotator is involved, in order to resolve the discrepancy where possible.

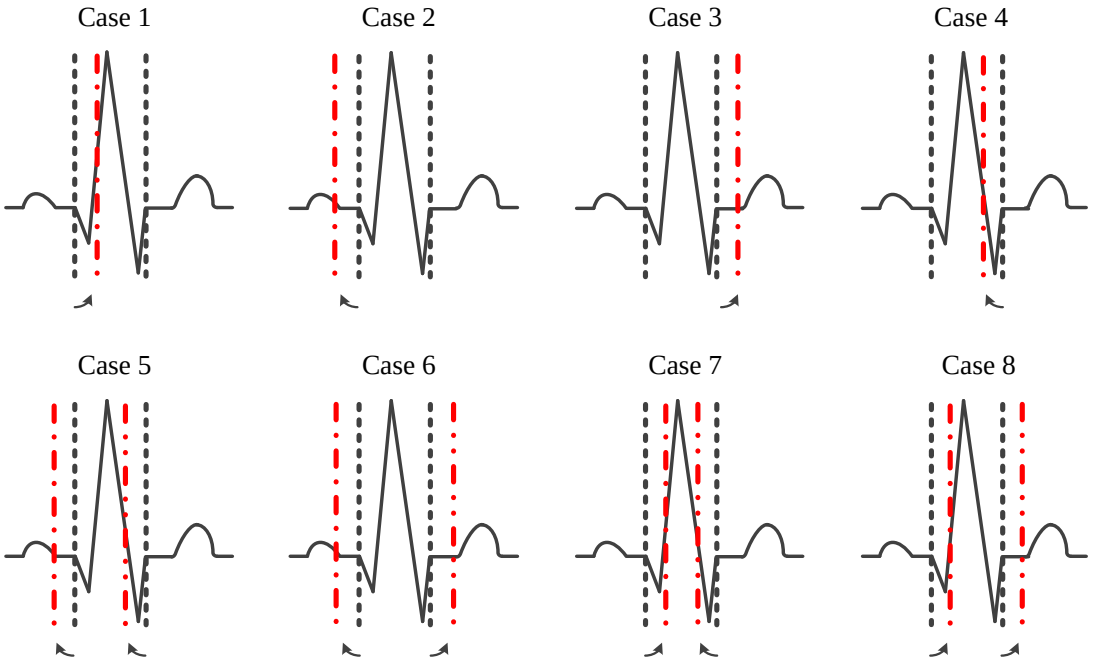


Figure 4.2: Artificial error injection that injects error to temporal location of wave boundary. Wave boundary of QRS is taken as example here. The injection covers eight possible cases of misdetection. Dashed line represents the correct wave boundaries as ground truth; dashed-dot and dash-dot-dot lines represent the misdetection biased by the artificial error on the onset and offset, respectively.

is introduced[5]. By taking into account these eight cases, the injection is able to provide a way of covering all possible scenarios of misdetection for the boundary detection of a wave component of interest. Therefore, by doing so in a systematic way, it may be possible to analyse the robustness of spectral energy.

4.4.2 Statistical Analysis of the Variation of Spectral Energy under Misdetection

Having introduced the way to inject artificial error, it is possible to investigate how the spectral energy would vary with respect to the changes of wave boundary detection. Increasing TIM introduces bigger misdetection error to the boundary detection of a wave component, hence the spectral energy of the corresponding wave. In this way the analysis of variations of spectral energy can then be made.

To start with, an isolated PQRST complex is considered from each record, as exactly what we did in Section 3.2.3.2. The human annotation of each wave component (P, PR, T, QRS and QT) of interest is provided by our cardiologist for each patient as the

[5]Since sampling frequency of our signals is 1 KHz, each millisecond represents one sample in data.

ground truth. As for the ground truth, TIM is equal to 0 with no artificial misdetection. Artificial error injection for these specific wave components is then applied separately based on their corresponding ground truths, thereby eight cases of misdetection can be derived.

Secondly, to examine whether spectral energy as a feature would exhibit consistency even when misdetection takes place, question may rise as in to what extent misdetection error should be injected. To answer this, we believe that it is sensible to think from the fiducial point detection algorithm point of view. This is because the misdetection error is primarily rooted in and generated by the algorithm. If spectral energy could address the misdetection issue and mange to show consistency, even given worst-case scenario is considered, then it is supposed to serve the ECG classification well as a feature. Hence, we choose to use the typical misdetection error generated by the algorithms in Chapter 3. Those errors were expressed using μ and σ for different fiducial points. As we just mentioned, worst-case scenario is considered in this study. Therefore, the possibly worst misdetection error we can get from TDMG and HFDA algorithm can be derived as the larger one by comparing the deviant values from their μ by σ amount. The way to obtain worst misdetection error is applied to each fiducial point case. To make the experiments easier from DWT analysis standpoint, the error is rounded to the nearest integer. By doing so, rounded temporal error injection in millisecond would not create fractional part of the temporal location of the wave boundary. This in turn would not jeopardise our DWT analysis for further spectral energy calculation. Note that most of our data used in this experiment is mainly excerpted from PTBDB. Therefore, we only consider Table 3.5 in Chapter 3 for our misdetection error derivation under the worst-case scenario. Here, Table 4.8 shows the typical worst-case misdetection error from our fiducial point detection algorithms. The sampling frequency is 1 KHz.

Table 4.8: Typical worst-case misdetection error from fiducial point detection algorithms after considering both TDMG and HFDA cases.

	P		**PR**		**T**		**QRS**		**QT**	
	On	**Off**	**On**	**Off**	**On**	**Off**	**On**	**Off**	**On**	**Off**
Error (ms)	19	24	19	15	23	32	10	11	10	32

As the artificial error injection is designed to affect the onset, offset and both onset and offset, the typical worst-case misdetection error for onset is then used in Case 1 and 2, while the typical worst-case misdetection error for offset is used in Case 3 and 4. The one used in Case 5 to 8 is selected as the larger one between the above two. Doing so allows Case 5 to 8 to investigate the worst possible scenario of misdetection when both boundaries are affected. Following, Table 4.9 gives the misdetection error for the corresponding cases that will act as TIM in our analysis.

Table 4.9: Worst-case misdetection error used for corresponding cases during artificial error injection.

	Error (ms)				
Case	**P**	**PR**	**T**	**QRS**	**QT**
1-2	19	19	23	10	10
3-4	24	15	32	11	32
5-8	24	19	32	11	32

Thirdly, the derivation of spectral energy takes place. Specifically, spectral energy of the wave component is derived using its wave boundary of ground truth (i.e. TIM equals to 0), as well as wave boundary affected by error after running artificial error injection (i.e. TIM equals to the worst-case misdetection error). Now, simple normalisation[6] with respect to spectral energy of ground truth must be carried out, which is applied to the spectral energy of ground truth itself as well as after injection. Doing so allows us to unify each patient's spectral energy of ground truth to be 1, and any spectral energy affected by error would simply exhibit a deviation from 1. Such deviation is referred to as **bias** in this study, and it is presented in percentage. It may be argued that normalisation may not be necessary. However, since spectral energy of a wave for one patient may be quite different to the other patient, normalisation with respect to ground truth erases the diversity, making it possible for fair investigation on consistency of spectral energy. Overall, doing so allows us to see how much variation of spectral energy we would encounter in each patient when worst-case misdetection happens. Applying this idea to all patients allows us to observe the deviation of spectral energy from a statistical standpoint. From there, it may show whether spectral energy as a feature would exhibit consistency. Note that the whole process of deriving spectral energy here is applied to seven spectral energy of our interest P_5, PR_5, T_5, QRS_2, QRS_3, QT_{35}, QT_{345}. Also, the same flow of analysis is generalised to 12 leads, allowing us to analyse the consistency of spectral energy in all available leads.

Fourthly, once the bias of each case from each patient in every lead is obtained, statistical analysis can be performed. In a particular lead, since we know the data of the bias of our entire patient database, the mean difference, or mean bias in our case, μ_{bias} and its standard deviation σ_{bias} can be calculated by

$$\mu_{bias} = \frac{\sum_i bias_i}{N} \tag{4.8}$$

$$\sigma_{bias} = \sqrt{\frac{\sum_i (bias_i - \mu_{bias})^2}{N}} \tag{4.9}$$

[6]It is a standard practice to scale the entire dataset (including training and testing dataset) to proper range prior to classification in machine learning [192].

where i denotes the index of patient and N denotes the total number of patients. Once we obtain the mean bias μ_{bias}, one-sample t-test [219] is performed with μ_{bias} as reference value[7]. Doing so also allows us to obtain the 95% confidence interval for μ_{bias} (here, by default we use 5% as the significance level).

In addition, together with μ_{bias} and σ_{bias} an important metric called *limits of agreement* can then be derived. The metric has been advocated by J.M Bland in one of the most well-known works in medical statistics [220]. In his work, he specifically stated a much reasonable approach, which involves *mean bias*, its *95% confidence interval* and *limits of agreement*, in clinical measurement to investigate whether a new measurement technique agrees sufficiently with an established one. The study particularly countered the notably misleading metric – *correlation coefficients*. Greatly inspired by this work, our study applies the same logic to evaluate the consistency of spectral energy. But instead of agreement for measurement, we regard our spectral energy of ground truth as the one we are going to compare to, and see whether agreement exists between such and spectral energy with worst-case misdetection error. To derive limits of agreement, we follow

$$\begin{aligned} limit_{up} &= \mu_{bias} + 2\sigma_{bias} \\ limit_{low} &= \mu_{bias} - 2\sigma_{bias} \end{aligned} \tag{4.10}$$

Here, upper limit and lower limit of agreement defines the range of limits of agreement. If the biases are normally distributed, 95% of the biases will be expected to lie within these limits. Since we have removed a lot of variation between patients, the bias is likely to follow a normal distribution. Bear in mind that, technically speaking the concept of 95% of the biases lying within the limits of agreement is different to 95% confidence interval for μ_{bias}, as the latter one mainly indicates 95% chances of having the mean of the bias (i.e. μ_{bias}) exist within the confidence interval. Overall, as exampled in [220] and [221], two measurement methods can be proved to agree to each other if the range of the limits is small enough within a clinical acceptable range. Similarly in our study, limits of agreement would indicate whether spectral energy affected by misdetection error is still consistent/agrees to the ground truth one. If it is small enough, it would statistically evidence our point of this study.

Finally, to illustrate what we have discussed so far, we take P_5 Case 1 as an example. The bias distribution in each lead, associated mean bias and 95% confidence interval can

[7]The idea of performing one-sample t-test is to explore the precision of the mean bias μ_{bias} in our study by hypothesis testing. Specifically, confidence interval is the measure of such precision, with 95% by default. Here, the null hypothesis is defined as sample mean bias μ_{bias} equals to the mean bias of the whole data distribution, and otherwise for alternative hypothesis. Following the standard exercise, confidence interval can then be derived and used in our later analysis.

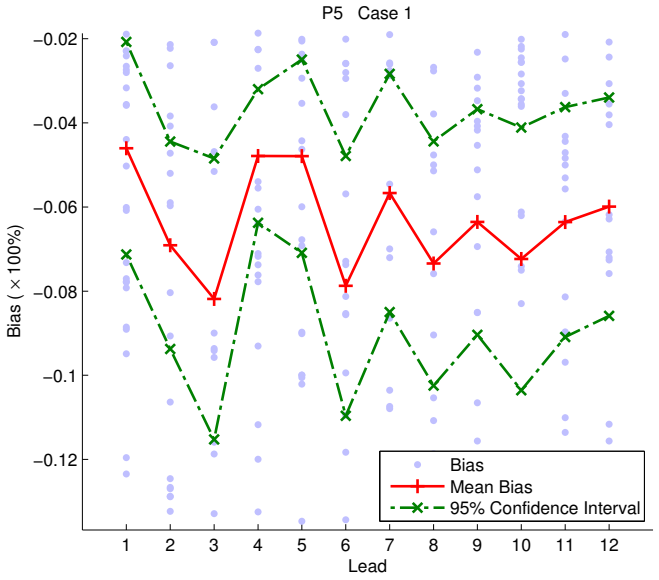


Figure 4.3: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of P_5 Case 1 (as an example in this study).

be seen[8] in Figure 4.3. As it shows, data of the bias is represented by dot. Depending on whether the worst-case misdetection error leads to gathering more spectral energy or losing it when calculating the spectral energy after misdetection, the bias will be presented as positive or negative value. The whole biases in each separate lead distribute in a vertical fashion (not horizontal) for the sake of plotting. In P_5 Case 1, generally the spectral energy of P wave under misdetection is smaller than the ground truth one, because the duration shrinks (Figure 4.2). Thus it shows negative y-axis index. In cases where the wave boundary is expanded (e.g. Case 2, 3 etc.), mostly the spectral energy would increase. Moreover, the entire figure depicts the bias distributions for all 12 leads. On top of it, mean bias and its 95% confidence interval of each lead is plotted. In addition, to give a fair comparison between cases taking into account all the leads, average of mean bias and its 95% confidence interval from the 12 leads is derived (upper and lower bound of the 95% confidence interval is done separately). Accordingly, the average of the limits of agreement for P_5 Case 1 is also derived. These three averaged metrics will be presented in table for discussion in next section (Section 4.4.2.1). Notice that figures of bias distribution, mean bias, 95% confidence interval of each lead and each case for different spectral energy of wave components of interest are covered in Appendix C.

Note that, it may be argued 95% confidence interval of a lead may only make sense when it is considered together with mean bias of that particular lead. Averaging 95%

[8]For the sake of compact plotting, the figure is presented in a zoomed-in view instead of a complete view of the bias distribution. Doing so allows us to focus only on where mean bias and 95% confidence interval show up, at the expense of having some data of bias not being seen in the scope, which does not matter in this study. It applies to the rest of the figures in the Appendix C as well.

confidence interval from 12 leads may not provide a sensible metric. Similar argument may also be raised when averaging the limits of agreement from 12 leads. However, since we treat each of the 12 leads equally, they are expected to follow similar distribution of bias (which is normal distribution in this study) and thus averaging, i.e. average of mean bias, 95% confidence interval and limits of agreement is practicable in this sense. Furthermore, as we will show later, apart from Case 1 same averaging policy applies to each case. Ultimately, with all 8 cases, a final averaged result is obtained. Again, it may also be argued that different cases may reveal totally different limits of agreement (as what we will see in Table 4.10 as an example), and averaging over 8 cases may not be appropriate. However, reason of doing so is not only because every case of misdetection has equal chance of happening and thus it make senses to average them from probability point of view, but also to give us an overall idea of how the averaged limits of agreement would exhibit for a particular type of spectral energy of our interest so that comparison with others can be made. Overall, the above description applies to other types of spectral energy of our interest as well. Following, discussion on the statistical analysis of consistency of spectral energy will be given in next section.

4.4.2.1 Experimental Results

In this section, the experimental results are obtained following the procedure we discussed in the last section. To analyse the results, we opt to consider separate frequency groups as we defined in Section 4.3.5.2.

First of all, Table 4.10 lists the experimental results, including averaged mean bias, 95% confidence interval, limits of agreement, of the low-frequency group (P_5, PR_5, T_5) of spectral energy of our interest. Discussion on this table is given in the following.

P_5: As can be seen, apart from Case 3, 6 and 8, generally the mean bias of P_5 shows relatively low value within ±23.0%, with confidence interval within ±28.0% in the rest of the cases. However, in Case 3, 6 and 8, mean bias lies at around 200.0% with confidence interval around -40.0% to 430.0%, making the overall mean bias and confidence interval fairly expanded and become bigger than the majority. This gives us the averaged level of P_5 to be 68.2% as mean bias and -22.9% to 159.4% as confidence interval. Thus, essentially it implies that P_5 under worst-case misdetection error tends to bias from the spectral energy of ground truth by between -22.9% and 159.4%. In the meantime, limits of agreement for P_5 is seen to exhibit similar phenomenon, as this metric is also mainly affected by the bias distribution as we encountered in deriving the former two metrics. For cases except 3, 6 and 8, the limits all locate within ±100.0%. However, when counting in these three cases, the averaged limits of agreement expand dramatically to -868.7% and 1005.2%. It makes us tend to believe that it is almost unacceptable to

agree that spectral energy under worst-case misdetection error is consistent with spectral energy of ground truth for P_5.

The reason of encountering severe variation of spectral energy in Case 3, 6 and 8 can be given as the following: referring back to Figure 4.2, it shows that these three cases all involve injecting misdetection error to the offset of P wave. This means they are expected to share similar effects in deriving spectral energy of P wave. As we have already mentioned in Section 4.3.5.2, spectral energy of P wave is solely drawn from the detail coefficients at DWT decomposition level 5. At level 5 the corresponding temporal resolution is 32 ms per coefficient. The worst-case misdetection error (Table 4.9) injected to the offset of P detection is 24 ms. Although the temporal resolution is 32 ms per coefficient, for some patients such 24 ms misdetection of the offset of P wave to the right could move boundary up to the edge of QRS complex. When translates into time-frequency domain, it in turn brings in the corresponding detail coefficient (specifically it would be one detail coefficient at level 5, because 24 ms can only cover maximum one coefficient at level 5) of QRS complex. Such one coefficient can ultimately increase the spectral energy a lot, which in this case causes a fairly huge variation to the spectral energy of ground truth.

PR_5: As for PR_5, a similar observation can also be found. Again, apart from Case 3, 6 and 8, mean bias of PR_5 shows relatively low value within ±22.0%, with confidence interval within ±28.0%. Nonetheless in Case 3, 6 and 8, this time it shows much more severe mean bias and confidence interval than P_5, where worst among the three – Case 6's mean bias goes up to 7238.0% and confidence interval spans from 2315.7% to 12160.4%. Taking into account these three cases, the averaged level mean bias and confidence interval of PR_5 lies at 2503.4% and 670.7% to 4336.1%, respectively. These huge numbers imply that, PR_5 under worst-case misdetection error tends to bias from the spectral energy of ground truth by between 670.7% and 4336.1%, which is of significant variation. On the other hand, limits of agreement follow similar pattern. Without Case 3, 6 and 8, averaged limits locates within ±90.0%. With Case 3, 6 and 8, the averaged limits enlarge significantly to -16344.6% and 21351.4%. Such huge span of limits prompt us to disagree to the fact that PR_5 under worst-case misdetection error is consistent with spectral energy of ground truth.

Table 4.10: Averaged mean bias, 95% confidence interval, limits of agreement for low frequency group (P_5, PR_5, T_5) from 12 leads for each of the eight cases, and the final overall average.

			Case								
	Metric (%)		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Averg.**
$\mathbf{P}_5$	Mean Bias		-6.3	4.6	197.5	-12.8	-6.4	203.9	-22.3	188.0	**68.2**
	95% Confidence	Lower	-9.0	0.3	-36.8	-17.0	-14.9	-31.5	-27.6	-46.3	**-22.9**
	Interval of Mean Bias	Upper	-3.7	8.8	431.8	-8.5	2.1	439.3	-17.0	422.2	**159.4**
	Limits of	Lower	-33.9	-38.9	-2212.0	-56.5	-94.2	-2217.0	-76.4	-2220.7	***-868.7***
	Agreement	Upper	21.2	48.0	2606.9	30.9	81.3	2624.7	31.8	2596.6	***1005.2***
$\mathbf{PR}_5$	Mean Bias		-4.4	2.4	5612.8	-16.4	-14.8	7238.0	-21.6	7231.2	**2503.4**
	95% Confidence	Lower	-6.4	0.5	816.9	-22.2	-21.1	2315.7	-27.6	2309.5	**670.7**
	Interval of Mean Bias	Upper	-2.4	4.3	10408.6	-10.6	-8.5	12160.4	-15.7	12153.0	**4336.1**
	Limits of	Lower	-25.1	-16.7	-43708.0	-76.3	-79.6	-43383.8	-82.8	-43384.5	***-16344.6***
	Agreement	Upper	16.3	21.6	54933.5	43.5	49.9	57859.9	39.5	57847.0	***21351.4***
$\mathbf{T}_5$	Mean Bias		-5.2	19.5	8.0	-14.1	7.9	30.1	-22.9	-0.7	**2.8**
	95% Confidence	Lower	-7.0	-10.9	0.7	-17.7	-25.2	-7.2	-27.3	-8.1	**-12.8**
	Interval of Mean Bias	Upper	-3.4	50.0	15.4	-10.5	41.0	67.3	-18.5	-6.6	**18.5**
	Limits of	Lower	-23.8	-293.6	-67.4	-51.1	-332.4	-353.0	-68.1	-76.0	***-158.2***
	Agreement	Upper	13.4	332.7	83.4	22.8	348.2	413.2	22.3	74.5	***163.8***

This time, having much severe values of the metrics in PR_5 than P_5 roots in the fact that the offset of PR with misdetection error inevitably runs into the zone of QRS complex. As we can see in the frequency response of Haar DWT (Figure 3.13), frequency bin of level 5 mainly covers frequency at around 30 Hz (given the fact that the sampling frequency f_S in our case is 1 KHz). According to the power spectrum of QRS complex, frequency components of this complex would also be reflected at this level in detail coefficient sequence. However, due to the very nature of QRS – short duration (normally 60ms to 110ms [46]) and more vitally, the disadvantage of having obscure time resolution at deeper levels, there are only 2 to 4 detail coefficients present at level 5[9]. Therefore, these coefficients can be disadvantageous when calculating spectral energy. Particularly it would seriously matter when misdetection occurs to the offset of PR interval. This is because the offset of PR is equivalent to the onset of QRS and therefore, misdetection would force the computation process of PR spectral energy to take into account the coefficients of QRS complex for most of the cases. This ultimately results into hugely severe variation of spectral energy.

T_5: Compared to P_5 and PR_5, T_5 has fairly consistent mean bias and 95% confidence interval throughout all cases, in the sense that no big difference between cases can be observed. The averaged values of both metrics are 2.8% and -12.8% to 18.5%, respectively. Hence it means misdetection error could lead to biasing from spectra energy of ground truth of T_5 by between -12.8% and 18.5%. On the other hand, limits of agreement for T_5 seem to be consistent within ±90.0%, except Case 2, 5 and 6. When these three cases are counted, the overall averaged limits of agreement lie at -158.2% and 163.8%.

Unlike P_5 and PR_5, T_5 shows much lower variation. Note that, even in Case 2, 5 and 6 where severe variation might be expected as misdetection error on T wave onset may absorb energy partly from QRS complex, they turned out to be of low extent of variation than severe cases of P_5 and PR_5 (Case 3, 6 and 8). This is mainly because the misdetection error for T wave is either 23 ms or 32 ms, for which translation from original time scale into level 5 gives us at most one detail coefficient. Remember that the preceding part (ST segment) and succeeding part (isoelectric baseline) of T wave devote to very low or even no frequency components, as we can normally observe in ECG signal. That means, this specific detail coefficient hardly contribute to high energy to the ground truth. Eventually, the energy with misdetection error tends to be similar to the ground truth and thus fairly low variation can be observed.

Having discussed the low-frequency group, we will present the high-frequency group in the following. Here, Table 4.11 shows the experimental results, including averaged mean

[9]Thus, characteristic coefficients of QRS complex are hardly useful at level 5. This is why we proceeded with level 3 in fiducial point detection for QRS (see Chapter 3).

bias, 95% confidence interval, limits of agreement, of the high-frequency group (QRS_2, QRS_3) of spectral energy of our interest. Note that, since QRS_2 and QRS_3 share the same complex and thus the same worst-case misdetection error on both onset and offset, we may carry out our discussion with them together.

QRS_2 & QRS_3: As can be observed, mean bias and confidence interval of both QRS_2 and QRS_3 show extremely low level of values in all 8 cases. Both shows zero averaged mean bias and nearly zero averaged interval. It strongly indicates that, the spectral energy of QRS complex at both levels exhibits consistency under worst-case misdetection. Regarding the limits of agreement, again fairly similar range throughout all 8 cases can be observed, making averaged limits to be -1.0% to 1.0% for QRS_2 and -0.7% to 0.7% for QRS_3. With such limits of agreement, it allows us to be confident that agreement between spectral energy with worst-case misdetection error and spectral energy of ground truth can be made for QRS_2 and QRS_3.

As we know, QRS complex mainly comprises high-frequency but short-duration wave. Normally 60 ms to 110 ms duration is observed and, when translating from original time scale into the scale at level 2 and level 3, we have 15 to 28 detail coefficients at level 2 and 8 to 14 at level 3, respectively. However, according to Table 4.9, the number of the coefficients that we may lose when calculating the spectral energy due to the worst-case misdetection error is 3 for level 2 and 2 for level 3, respectively. So, even if worst-case misdetection happens on both side of the wave, we still have 9 to 22 coefficients left at level 2 for QRS_2 and 4 to 10 coefficients left at level 3 for QRS_3. And bear in mind that, the major frequency components of QRS are mainly located in the centre of the complex, indicating that detail coefficients in the centre must contribute to the most of the spectral energy. Therefore, spectral energy of QRS complex with worst-case misdetection error would not vary much.

Again in Table 4.11, we have the experimental results of the combined frequency group (QT_{35}, QRS_{345}) of spectral energy of our interest. Since QT_{35} and QT_{345} share the same complex and thus the same worst-case misdetection error on both onset and offset, discussion on both of them is made together.

QT_{35} & QT_{345}: Similar to QRS_2 and QRS_3, mean bias and confidence interval of both QT_{35}, QT_{345} reveal extremely low level of values in all 8 cases. The average of both is further derived as -0.5% and -0.8% to -0.2% for QT_{35}, and -0.5% and -0.8% to -0.3% for QT_{345}, respectively. It means both spectral energies exhibit consistency under worst-case misdetection. Regarding the limits of agreement, very small interval of the limits can be found, rendering the averaged limits to be -3.4% to 2.4% for QT_{35} and -3.0% to 1.9% for QT_{345}. From there, we are confident

to agree that spectral energy for QT_{35} and QT_{345} are consistent as their energy with worst-case misdetection error does not deviate much from spectral energy of ground truth.

As we know, the major morphological patterns in QT complex are QRS complex and T wave. To obtain the spectral energy of this kind, level 3 and level 5 are good choices as they can mostly reflect the frequency components of the QT complex at these two levels, respectively. However, the transition of spectral energy between both levels is also worth considering. Though it may not be obvious to observe in time domain, the transition from QRS to T (i.e. high to low frequency) may devote to certain information. As a result, level 4 is taken into account in QT_{345}.

In addition, the way that spectral energy of QT is derived is different to the low- and high-frequency group. Summation of squared detail coefficients over multi-level is needed in this case. That actually leads to two concerns. Firstly, according to the results in Table 4.11, the inclusion of level 4 in QT_{345} does not make many differences to QT_{35}. This implies that the transition of the energy barely devotes to the final energy. Secondly, the worst-case misdetection error is 10 ms for onset and 32 ms for offset (Table 4.9), respectively. Translating them into number of coefficients at different decomposition levels give us at most 2, 1 and 1 coefficient for onset, and at most 4, 2, 1 coefficient for offset at level 3, 4 and 5, respectively. From there, the loss or the gain of coefficient (depending on which case of misdetection) due to the misdetection when calculating the spectral energy would not matter too much, in particular level 5, where the majority of the spectral energy of the complex lies. That is because, for the onset, it is very rare to miss or gain 1 coefficient at level 5 as only 10 ms is misdetected at the original time scale. For the offset, it is deemed to miss or gain 1 coefficient. But since it takes place on T wave, T wave barely devotes to the total energy compared to QRS complex. As a result , no matter onset or offset, misdetection on QT complex does not create much variation to the spectral energy of ground truth.

4.4.2.2 Discussion

Overall, putting together the averaged metrics of low-, high- and combined frequency group, key observations can be made as follows.

Firstly, averaged mean bias and its associated 95% confidence interval are discussed here. Clearly, from what can be seen in Table 4.10 and Table 4.11, low-frequency group tends to show high mean bias and the confidence interval is also large (in particular PR_5), whereas high- and combined frequency groups exhibit extremely low value in both metrics. This actually implies that spectral energy of low-frequency group with worst-case misdetection error tends to bias from the ground truth to more severe extent than the other two groups.

Next, as can be seen that, limits of agreement for the three groups reveal similar phenomenon. As suggested in [220], limits of agreement is the key judge. We further apply it and judge whether spectral energy with misdetection error and spectral energy of ground truth is agreed (in our case, consistent) or not in accordance with certain satisfactory agreement. Though spectral energy as a feature is used in this thesis and we find it promising, as far as we concern there is no clinically-proven satisfactory agreement to judge its good and bad. Speaking of satisfactory agreement, [222] explicitly states that, in medical measurement, how far apart measurements between two different methods can be without being problematic depends on the use to which the result is put, and it is a question of clinical judgement. Putting the same logic into our context, that means boundary of satisfactory agreement should depend on what we are going to use spectral energy for. In fact, in our work, spectral energy is used in classifying normal and abnormal ECG. Thus, it has to be coupled with classification algorithm to draw deeper conclusion, which will be discussed in the next section. But for now, our main goal is to statistically estimate how such variation would take place under misdetection and summarise whether it is acceptable or not.

To do that, we choose to use ±5% as our satisfactory agreement[10] to make the judgement. As for low-frequency group, more than ±160% limits is observed in all three spectral energy. This is totally unacceptable, as it tells that the range is so large that 95% of subjects deviate from the spectral energy of ground truth by the extent of ±160% limits, let alone with more severe situations, like P_5 and PR_5. On the other hand, high-frequency group exhibits not larger than ±1% limits, while combined frequency group exhibits not larger than ±3.4% limits. From there, we can see that the limits are within ±5%, and it effectively indicates 95% of subjects deviate from the spectral energy of ground truth by an acceptable extent.

[10]Similar to the idea of 5% significance level in hypothesis testing, the null hypothesis here can only be not rejected when more than 95% of the observations (i.e. biases) are not beyond the limits of agreement. If we set the satisfactory agreement to 5%, that means as long as the limits of agreement of one spectral energy is less than 5%, the majority of the observations would lie within 5% (assuming that it is Gaussian). Thus, it can be safely stated that this spectral energy is consistent.

Table 4.11: Averaged mean bias, 95% confidence interval, limits of agreement for high-frequency group (QRS_2, QRS_3) and combined frequency group (QT_{35}, QT_{345}) from 12 leads for each of the eight cases, and the final overall average.

			Case								
	Metric (%)		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Averg.**
QRS_2	Mean Bias		-0.2	0.1	0.2	-0.2	-0.1	0.3	-0.3	0	**0**
	95% Confidence	Lower	-0.2	0	0.1	-0.3	-0.2	0.1	-0.5	-0.1	**-0.1**
	Interval of Mean Bias	Upper	-0.1	0.2	0.2	-0.1	0	0.4	-0.2	0.1	**0.1**
	Limits of	Lower	-0.9	-0.5	-0.7	-1.2	-1.1	-1.1	-1.9	-0.9	***-1.0***
	Agreement	Upper	0.6	0.7	1.1	0.9	0.9	1.7	1.2	0.8	***1.0***
QRS_3	Mean Bias		-0.1	0.1	0.1	-0.1	0	0.1	-0.2	0	**0**
	95% Confidence	Lower	-0.1	0	0.0	-0.2	-0.1	0	-0.3	0	**-0.1**
	Interval of Mean Bias	Upper	0	0.1	0.1	0	0	0.2	0.0	0.1	**0.1**
	Limits of	Lower	-0.3	-0.5	-0.5	-1.1	-0.8	-0.8	-1.3	-0.6	***-0.7***
	Agreement	Upper	0.2	0.6	0.6	0.9	0.7	1.1	1.0	0.6	***0.7***
QT_{35}	Mean Bias		0	0	0.1	-0.4	-0.2	0.4	-2.2	-1.7	**-0.5**
	95% Confidence	Lower	-0.1	0	0	-0.6	-0.4	0.2	-3.0	-2.4	**-0.8**
	Interval of Mean Bias	Upper	0	0.1	0.2	-0.3	0.1	0.6	-1.4	-0.9	**-0.2**
	Limits of	Lower	-0.5	-0.3	-0.6	-1.9	-2.5	-1.8	-10.4	-9.5	***-3.4***
	Agreement	Upper	0.4	0.3	0.9	1.0	2.2	2.6	6.0	6.1	***2.4***
QT_{345}	Mean Bias		0	0	0.1	-0.4	-0.2	0.3	-2.3	-1.9	**-0.5**
	95% Confidence	Lower	-0.1	0	0	-0.5	-0.3	0.2	-3.1	-2.6	**-0.8**
	Interval of Mean Bias	Upper	0	0	0.1	-0.3	0	0.4	-1.6	-1.1	**-0.3**
	Limits of	Lower	-0.3	-0.2	-0.5	-1.4	-1.6	-1.0	-9.9	-9.3	***-3.0***
	Agreement	Upper	0.2	0.2	0.6	0.7	1.3	1.6	5.3	5.5	***1.9***

4.4.3 Classification Performance using Spectral Energy as a Feature under Misdetection

As what we have seen from Section 4.4.2, spectral energy can be viewed as a consistent feature particularly in high-frequency and combined frequency group. Since our main goal is to utilise spectral energy to serve classification, such consistency may further influence the ECG classification. On this basis, it raises a concern of how to examine to what extent the consistency of spectral energy against misdetection would affect the classification. This essentially leads us to the following exploration – classification performance affected by the misdetection error. In this case, artificial error injection (Figure 4.2) may be deployed to couple with classification models to accomplish the task.

4.4.3.1 Jitter Effect

In classification, each sample can be represented as an individual vector of features in the feature space. Any changes to the vector would bias it from the original one. Similarly, it may be argued that misdetection in ECG feature detection algorithm would bias the vector of spectral energy features with certain offset. This type of offset can be considered as *jitter*. The notion of jitter[11] in this work is referred as offsets to the original sample, which would eventually influence the classification performance. Here Figure 4.4 shows the conceptual demonstration of jitter effects in a 2-dimentional feature space, which is constructed by dimension 1 and 2. As we can see in Figure 4.4 (a), with decision boundary, it is able to classify most of class 1 (red circle) and class 2 (blue triangle) successfully. Only two of class 1 (green circle) and three of class 2 (purple triangle) are misclassified. But what if error is introduced to dimension 1 of every data point? In this case, all data points are biased by the extent of error, as shown in Figure 4.4 (b). Therefore, such phenomenon pushes those of class 1 on the edge of *classified* to be *misclassified* while those of *misclassified* class 2 back to *classified*. In the end, it actually changes the end results of classification. Also bear in mind that, as a generic and ideal example here, error on dimension 1 applies to every record to the same extent of bias. However, most probably it may not be the case in real scenario, since every data point (i.e. sample) has its dedicated way of getting error in every possible dimension during the feature detection stage preceding the classification stage. The intention of this conceptual jitter effects is to show how error could bias the feature vector and lead to unexpected results.

[11] To avoid latent confusion with our work, in machine learning community the term *jittering* is generally referred as a technique that introduces random offsets to the data help observe the data clearly [223].

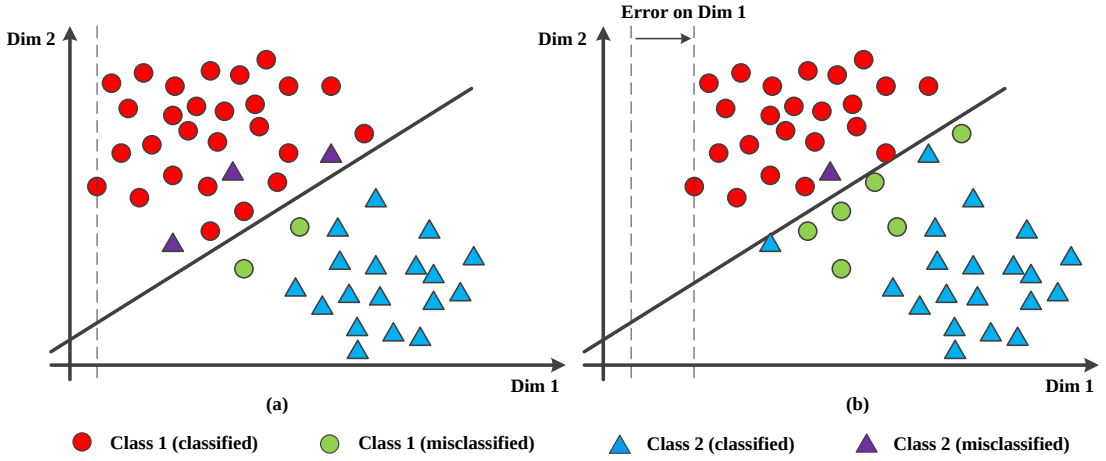


Figure 4.4: Conceptual demonstration of jitter effects. (a) the original 2-D feature space; (b) feature space with dimension 1 biased by error.

4.4.3.2 Modified 10-fold Cross-Validation for Jitter Effect

In general, in order to evaluate the predictive performance of the classification model, we have a separate validation set to perform validation. When the dataset is very small, therefore small validation set, the performance derived could be problematic [142]. That is why we use k-fold CV technique, as it allows us to make more efficient use of the limited data we have. Furthermore, it has also been proved that stratified 10-fold CV was recommended for model selection [194]. However, modification on 10-fold CV is needed to form a proper method to evaluate jitter-based performance for classification models in our study.

As we will see, this modified 10-fold CV will be applied to not just the scenario where TIM equals to WCE, but also those less than that. Previously in Section 4.4.2, focusing on worst-case scenario is sufficient as we only need to investigate the distribution of spectral energy under extreme scenario without any concerns of classification. Doing so automatically covers any possible cases of misdetection within worst-case scenario. However, for classification it is different. Varying spectral energy could easily affect the classification performance. To investigate the robustness of spectral energy as a feature in this context, we need to consider every possible case of misdetection, i.e. TIM equals 0 up to the worst-case scenario error. Despite saying so, if we follow the conventional way of cross validation, it would otherwise be pointless as the training dataset would be shuffled every time when we run 10-fold CV to evaluate new TIM. Thus, the averaged results would not make sense in the end. So, changes have to be made in the way that every TIM ranging from 0 to WCE from a specific validation dataset can be tested separately on the same trained model during the cross validation process. To our best knowledge, there is no specific analysis method that is dedicated to jitter effects on

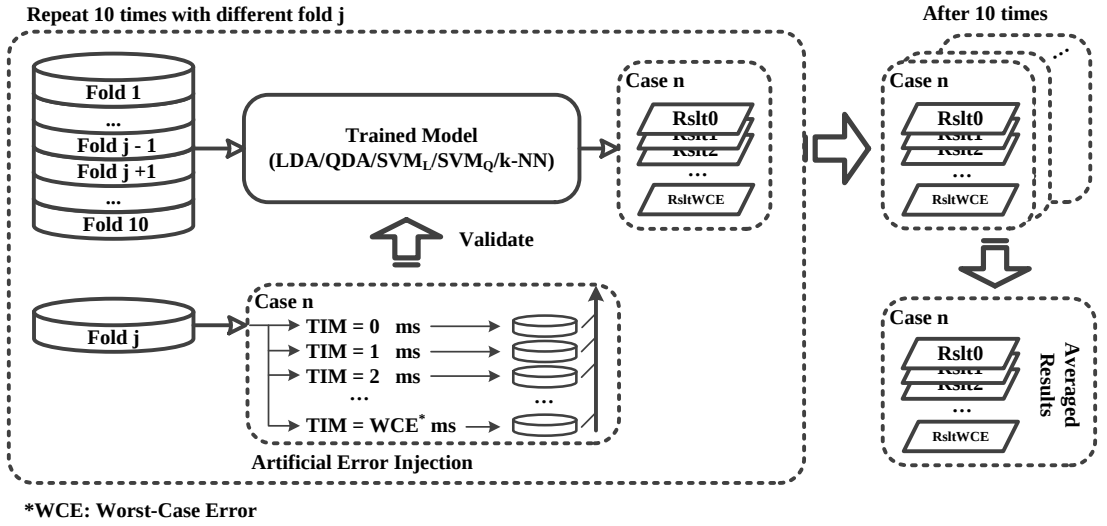


Figure 4.5: The proposed modified 10-fold CV for jitter effect.

classification performance with 10-fold CV. As a result, we have proposed a modified 10-fold CV that utilises the fundamental theory of k-fold CV.

An overview of modified 10-fold CV is illustrated in Figure 4.5. Specifically, it is a process where 9 folds of the samples are used to train the classification model, while artificial error injection (Figure 4.2) is applied to the validation fold j in order to construct misdetection scenarios for Case n, where $n = 1 \sim 8$. The trained model is then validated by each scenario in turn and further produces the corresponding classification results. Secondly, the whole process repeats 10 times to produce 10 sets of classification results, where each fold acts as validation set in each time. Here, shifting the validation set from one fold to another allows jitter to inject into each validation fold, eventually spread out to all PQRST complexes. That enables the thorough study of jitter effect across the whole dataset. The final results can then be obtained by averaging over these 10 sets. By then, we have the averaged result of each scenario for Case n. Notably, the fundamental part of this method is to have validation fold subjected to the artificial error injection, which means all misdetection scenarios have the same trained model to validate upon. In this way, we not only set up a reasonable trained model (as we normally have in normal 10-fold CV), but also inject errors that are of our interest to the validation data. This helps achieve reasonable results for jitter effect in classification. Note that ultimately the jitter is applied to

As usual, the classification algorithms is coupled with the training set to produce a trained model to be tested by the validation set with artificial error injection. However, some particular concerns of our proposed method regarding classification part should be discussed, and they are listed as follows.

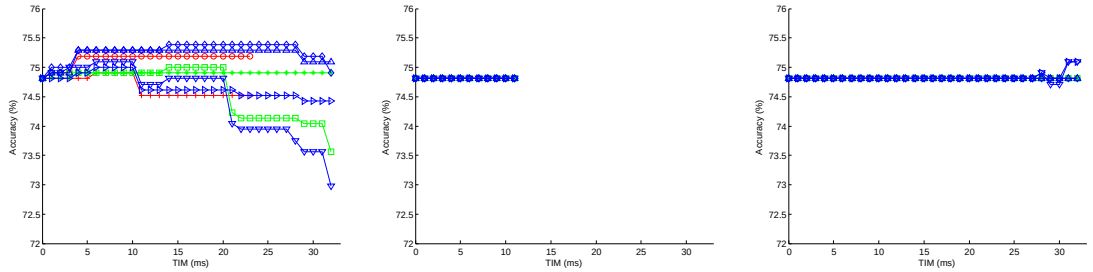
- In normal 10-fold CV, classification algorithm is usually trained on the training set and tested on the validation set. Intuitively, it might make sense in our study as well if we applied artificial error injection to both training and validation set and obtained classification result under a specific TIM, and from there we moved on to the other TIMs. However, we cannot directly apply such strategy. This is because the main goal of this study is to observe how jitter effect would affect classification performance based on the ground-truth data (i.e. spectral energy of ground truth). So, we need to have trained classification model based on ground truth, and derived only from training dataset. Given such, we can test it with validation set injected with artificial error. As a result, artificial error injection cannot be simply deployed to both training and validation set, but validation set only.

- Except k-NN, the rest of classification algorithms are parametric and should be optimised based on the data we have. For LDA and QDA, the class prior probabilities are estimated from the class relative frequencies in training set. Since we have 52 normal versus 52 abnormal samples and also the way we perform modified 10-fold CV follows a fair division of the classes, we would have the same prior probability (50% versus 50%) for both classes. On the other hand, generalisation parameter C for both SVM_L and SVM_Q should be optimal in terms of accuracy and computational complexity. Generally it should be chosen to achieve the best generalisation capability and smallest number of support vectors [185]. So in this study, we consider training set (without validation set fold j) as an independent dataset and use it to run a separate 10-fold CV specifically for searching the optimal generalisation parameter C. Here, the search ranges from 2^{-5} to 2^{10}.

- To fully utilise the entire dataset and check the generalisation of the classifiers for jitter effect, 10 runs of modified 10-fold CV is performed.

Note that before we come to the next section and examine the jitter effect, the use of lead scenario based optimal spectral energy sets in Table 4.6 must be mentioned here. As before, these optimal sets are mainly deployed as the feature vectors in our exploration. However, it may be argued that our study of jitter effect is not applicable and cannot be generalised to broader situation where no optimal feature set as such is pre-defined. But since our work is the first preliminary study of this kind, we attempt to examine how spectral energy would affect classification performance, even with the best optimal feature set, to see whether spectral energy is a reasonably consistent feature.

Now, as we aim to examine spectral energy as a feature, it is reasonable to observe variation of only one feature at a time. Hence, one selected spectral energy feature is subjected to modified 10-fold CV. But if there are more than one feature in the optimal feature set, the rest of the features are kept constant, which means they are unbiased (i.e. TIM = 0). In this way, the rest of the features would not affect that particular



(a) Feature 1: T_5 (b) Feature 2: QRS_2 (c) Feature 3: QT_{35}

Figure 4.6: Examples of variation of classification performance for 1 lead scenario for LDA, where Case 1: '+', Case 2: '$\circ$', Case 3: '$*$', Case 4: '$\square$', Case 5: '$\diamond$', Case 6: '$\triangle$', Case 7: '$\triangledown$', Case 8: '$\triangleright$'.

feature under investigation for misdetection error in terms of classification performance. In addition, to evaluate the performance of each of the classifiers we use the metrics of overall testing accuracy, as defined in Equation 2.39.

4.4.3.3 Results and Discussion

To illustrate the result based on the scheme, let us take 1 lead scenario from LDA for example. This feature vector consists of 3 features from Lead 3, namely, T_5, QRS_2, QT_{35} (see Table 4.6). By applying modified 10-fold CV to each feature at a time, Acc at each TIM (0 to respective WCE) for all 8 cases of misdetection can be obtained, as shown in Figure 4.6. Notice that all cases of Feature 1 (T_5) except Case 3 tend to slightly fluctuate along the TIM to different extents. Unlike Feature 1, Feature 2[12] exhibits fairly constant trend up to its WCE, and so does Feature 3 except tiny fluctuation from Case 7 and 8 at the end. In fact, these are expected as low-frequency group tends to show inconsistency while high- and combined frequency groups do not (see Section 4.4.2.2).

With this example in mind, observation can be generalised to other feature vectors of optimal spectral energy sets based on lead scenario. Ultimately, every spectral energy feature in each of 5 lead scenarios for each classifier has gone through modified 10-fold CV individually, and finally obtained the results in accordance with the example above. To effectively analyse the variation using these data on a unified platform, we proposed a metric called Variation of Classification Performance (VCP). With VCP, it is able to evaluate the fluctuation of a particular feature in all 8 cases as a whole. In fact, this metric utilises the concept of Frobenius Norm (or Hilbert-Schmidt Norm) [224], which essentially calculate the $N \times M$ norm $\sqrt{\sum_i^N \sum_j^M |X_{ij}|^2}$ (in our case, N = 8 and M = WCE). To further build upon it for our purpose, a penalty weight t is needed so that

[12]Reason of having accuracy data up to TIM = 10 to 11 ms in Figure 4.6(b), unlike the other two features, is because the worst-case misdetection error it can get is 10 ms in Case 1 to 2 as well as 11 ms in Case 3 to 8, in accordance with Table 4.9.

the higher variation at bigger TIM is, the larger VCP will be. In the end, VCP is defined as

$$VCP(X_{c,t}) = \frac{\sqrt{\sum_{c=1}^{8}\sum_{t=1}^{WCE}|(X_{c,t} - X_{c,0}) \times 100|^2 \cdot t}}{K} \tag{4.11}$$

where c and t denote case of misdetection and TIM, while $X_{c,t}$ and K denote the accuracy at certain c and t as well as normalisation factor, respectively. Here K equals to the summation of WCEs from 8 cases. Note that each $X_{c,t}$ is subtracted from $X_{c,0}$ so that $X_{c,t}$ is normalised with respect to the accuracy at TIM = 0. In this way the final score will not be biased to individual lead scenario. Overall, VCP is able to condense the performance of variation from all cases of misdetection. This is because each data point (e.g. in Figure 4.6) represents one independent classification performance at one specific TIM. Combining all the data points in the form of VCP would not conceal any implicit data distribution attribute. Thus, the lower the VCP is, the more consistent the feature under consideration can achieve in classification. Also, since VCP is a measure of consistency in terms of classification performance, it is only used to reflect the consistency against misdetection, not superiority in classification (e.g. lower/higher accurate rate).

Now, to author's best knowledge, so far there is no specific work on investigation of feature consistency targeting at normal and abnormal ECG classification. So, to truly justify spectral energy as a feature could perform consistently, comparison with other typical features in ECG classification should be made. Here we opt to use Wave Duration (WD) of the five wave components, namely P, PR, QRS, QT, T[13] as feature vectors in ECG classification, since WD are closely related to misdetection and therefore sensible comparison with spectral energy can be reached. After that, likewise what we did for spectral energy, each of them runs through modified 10-fold CV and eventually we obtain the VCPs for WD. However, note that in this case artificial error injection (Figure 4.2) does not apply to WD. This is due to the fact that there are only two sources of offsets that can bias the value of WD – either expansion or contraction of the duration. Therefore we only need to consider these two cases of misdetection for WD. The amount of WCE that would bias WD to worst-case expansion and contraction uses values of Case 5-8 in Table 4.9. But the only difference is, these values correspond to onesided misdetection whereas expansion and contraction of duration has two-sided misdetection. So, WCE for twosided misdetection has to be doubled. That means, for WD the WCE will be 24×2, 19×2, 32×2, 11×2, 32×2 for P, PR, T, QRS, QT respectively.

Table 4.12 lists the VCPs of both spectral energy and WD for the five classifiers, with each VCP indicating the consistency of individual feature against misdetection. Clearly, depending on the lead scenario and the classifier, features that are available in Table

[13]These five features will be denoted as Feature 1, 2, 3, 4 and 5 in Table 4.12 for display purpose.

4.6 has VCP and those not applicable in Table 4.6 are simply ignored in Table 4.12. Now, starting from LDA, it can be seen that for most of the cases, VCPs of features ranging from 1 lead scenario up to 5 lead scenario of spectral energy exhibit fairly low values, mostly close to 0. In comparison to that, VCPs of WD show extremely high values, all of which are 1 to 3 order of magnitudes higher than spectral energy's ones. Not only LDA, similar phenomena can also be observed in QDA, SVM_L, SVM_Q and k-NN. On one hand, it effectively tells us that, any feature involved in any lead scenario based on spectral energy classification would exhibit extremely consistent classification accuracy in comparison to WD, under all situations of misdetection cases with all possible misdetection errors ranging from ground truth up to worst-case misdetection. More importantly, it not only applies to specific spectral energy features under specific classifiers, but all of the features (low-, high- and combined frequency groups) and classifiers. That means spectral energy as a feature does show the potential of consistency against misdetection with varied type of classifiers in application like classification. On the other hand, WD show relatively inferior inconsistency against worst-case misdetection. None of WD feature shows VCP similar to spectral energy feature under corresponding classifier.

In addition, if we take a deeper look at the table together[14] with Table 4.6, we can see that VCPs of low-frequency group features in general exhibit roughly 1 order of magnitude higher than high- and combined frequency groups. This essentially tells us that, the classification performance of low-frequency group is more likely to fluctuate when misdetection occurs, which in turn indicate that this group is less consistent than high- and combined frequency group. Notably, such observation should be expected as what we concluded in Section 4.4.2.2 should be reflected in classification phase, and the whole in this study thus further justify the previous conclusion.

4.5 Concluding Remarks

In this chapter, we have presented investigations into spectral energy of an ECG complex, as well as the robustness of spectral energy against misdetection from two different perspectives: statistical analysis of the variation of spectral energy and classification performance using spectral energy as a feature, both under worst-case misdetection.

Firstly, we attempted to explore four different approaches of deriving spectral energy of an ECG complex. Associated four sets of experiments were conducted, namely DFT on entire complex, DWT on entire complex, adaptive thresholding policy on entire complex, and finally DWT on specific wave components of the complex. The results have shown that, DWT-based spectral energy on specific wave components shows fairly better

[14] What it means is, since VCPs in Table 4.12 are arranged according to the participating features in certain lead scenario under certain classifier, it is possible to link the VCP to spectral energy feature that is of low-frequency, high-frequency, or combined feature group as shown in Table 4.6.

Table 4.12: VCP of 5 lead scenarios for spectral energy as well as of WD for all classifiers.

	Fea.	**VCP of a specific feature under misdetection**									
	Space	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
LDA	**1 LS**	0.13	0.00	0.02	/	/	/	/	/	/	/
	2 LS	0.06	0.00	0.18	0.02	/	/	/	/	/	/
	3 LS	0.23	0.00	0.01	0.14	0.02	0.02	0.02	0.00	/	/
	4 LS	0.18	0.00	0.01	0.18	0.00	0.01	0.01	0.01	0.03	/
	5 LS	0.00	0.21	0.00	0.14	0.00	0.00	0.00	0.10	/	/
	WD	18.29	7.65	36.49	42.10	8.73	/	/	/	/	/
QDA	**1 LS**	0.24	0.01	/	/	/	/	/	/	/	/
	2 LS	0.11	0.01	0.02	0.11	0.00	0.01	/	/	/	/
	3 LS	0.12	0.01	0.08	0.01	0.07	0.01	0.03	/	/	/
	4 LS	0.10	0.01	0.00	0.04	0.01	0.05	0.00	0.04	0.00	/
	5 LS	0.16	0.01	0.02	0.02	0.01	0.03	0.09	0.04	0.00	/
	WD	49.80	13.26	59.14	28.99	7.78	/	/	/	/	/
SVM$_L$	**1 LS**	0.18	0.00	/	/	/	/	/	/	/	/
	2 LS	0.13	0.15	0.00	0.02	/	/	/	/	/	/
	3 LS	0.17	0.17	0.02	0.03	0.02	0.03	/	/	/	/
	4 LS	0.18	0.19	0.02	1.67	0.00	0.01	0.11	0.01	/	/
	5 LS	0.08	0.18	0.00	0.00	0.03	0.00	0.01	0.16	0.00	0.01
	WD	4.52	3.56	25.85	34.15	3.80	/	/	/	/	/
SVM$_Q$	**1 LS**	0.29	0.00	/	/	/	/	/	/	/	/
	2 LS	0.29	0.21	0.00	0.00	/	/	/	/	/	/
	3 LS	0.08	0.07	0.00	0.02	0.07	0.00	0.01	/	/	/
	4 LS	0.14	0.00	0.27	0.06	0.23	0.01	0.03	/	/	/
	5 LS	0.04	0.07	0.00	0.08	0.06	0.20	0.00	0.01	/	/
	WD	16.71	15.41	55.43	5.81	19.73	/	/	/	/	/
k-NN	**1 LS**	1.30	/	/	/	/	/	/	/	/	/
	2 LS	0.19	0.29	/	/	/	/	/	/	/	/
	3 LS	0.00	0.00	0.09	0.00	0.00	0.03	0.00	/	/	/
	4 LS	0.00	0.00	0.02	0.00	0.00	0.04	0.00	0.04	/	/
	5 LS	0.00	0.00	0.01	0.00	0.05	0.00	0.00	0.00	0.07	0.00
	WD	20.56	18.45	36.61	10.44	11.41	/	/	/	/	/

classification performance than the other three approach, with roughly 2-12% increase in each classifier. Though fiducial point detection algorithm is required for each wave component as an extra preceding step, the results have significantly stressed on our argument that it is worth to apply the fiducial point detection algorithm. Otherwise it would have been of no use to apply other approaches of spectral energy derivation that exhibit low classification accuracy, as it would not satisfy any clinical usage in our later studies.

Secondly, we also attempted to explore the performance of spectral energy as a feature in ECG classification. Two experiments were conducted – one was the statistical analysis

of the variation of spectral energy under misdetection, the other was classification performance assessment using spectral energy as a feature under misdetection. Regarding the first one, misdetection error was artificially introduced to the boundary localisation of our wave components of interest in a systematic way. From there, the consistency of spectral energy was examined given the worst-case error of detection of ECG feature extraction algorithm. Regarding the second one, classification performance of spectral energy variation along the temporal index of misdetection up to the worst-case error were also investigated, with dedicated five classifiers to explore the consistency spectral energy might exhibit. A scheme that is capable of examining jitter effect in classification, as well as a new metric for evaluating the variation were proposed. Overall, from the findings of both experiments, we concluded the followings: in general, low-frequency group shows unacceptable limits of agreement, indicating that it is almost deemed to deviate from the spectral energy of ground truth by a huge extent, in particular P_5 and PR_5, when worst-case misdetection error takes place. On the other hand, high- and combined frequency group show very promising limits of agreement. This renders them as being consistent even with worst-case misdetection error. In our second experiment, VCPs of spectral energy exhibited approximately 1 to 3 order of magnitude higher than wave durations'. That means, in general spectral energy exhibits supervisor consistency in terms of classification accuracy compared with wave duration, both of which are bound to be affected by misdetection. What is more, within the category of spectral energy, low-frequency group actually showed slightly inferior consistency than high- and combined frequency group. And this primarily roots in their responses to the deviation from spectral energy of ground truth, meaning that the difference between the former group and the latter two groups in consistency in the first experiment in fact affects the outcome of the second as a result.

Finally, we have seen that spectral energy as a feature has shown consistency against misdetection and achieved certain level of classification accuracy. But space for improvement still exists, partly because low-frequency group may not offer comparatively good class-relevant features than the other two groups. Therefore, in the next chapter we will explore whether some additional features could supplement spectral energy to enhance the overall classification performance.

Chapter 5

More Features to Enhance Spectral Energy-based Classification

As what we have concluded in Chapter 4, low-frequency group of spectral energy was deemed to show unacceptable consistency under misdetection, and also demonstrated that such inconsistency impairs the classification performance, particularly with LDA, SVM_L and SVM_Q. This in turn prompts us to believe what we have in Set 4 and its associated classification results (Table 4.7) may exist rooms for improvement. Many ways of achieving improvement are possible, for instance deploying more class-relevant features, advanced classification models, etc. In our study, we opt to explore and utilise potentially useful features to augment spectral energy upon classification. By adding more relevant and non-redundant features (Section 2.3.3.2), classification performance is expected to improve. Both classifications based on spectral energy with and without more features will be justified in single heartbeat scenario and multiple heartbeat scenario. In addition, bear in mind that computational complexity is a big concern in our study. So computational complexity required to label a new sample will be taken into account in this chapter. Coupling computational complexity with classification performance during analysis will direct us to a good balance between the two, and thereby derive a proper solution in implementing hardware for ECG classification in mobile environment in next chapter.

The rest of this chapter is organised as follows[1]. Section 5.1 covers the computational complexity required to label a new sample for trained classifiers. With it we will be able to judge the equivalent energy consumption classification may cost. Besides, two

[1]The contents of Section 5.1 to 5.3 have partly appeared as "Design of a Low-Power On-Body ECG Classifier for Remote Cardiovascular Monitoring Systems" by Chen *et. al.* in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*.

scenarios of heartbeat classification are considered: single and multiple, and briefly discussed in Section 5.2. As will be discussed in the following sections, multiple heartbeat classification effectively works as an extended study of the single one. Next, Section 5.3 discusses the spectral energy-based classification without any augment of features, specifically covering both scenarios of heartbeat classifications and parametrical tuning for optimisation of SVM. In Section 5.4, spectral energy-based classification with augment of features is covered, with detailed discussion on potentially useful features, feature selection algorithms and both scenarios of heartbeat classifications. At last, Section 5.5 concludes this chapter.

5.1 Computational Complexity of Our Selected Classifiers

Covered in Section 1.2, balance between accuracy and computational complexity is considered to be challenging work in mobile CVD monitoring systems. Therefore, it is necessary to investigate the fundamental complexity formulation for each of the five classifiers. Having discussed the basics of the five classifiers in Section 2.3.4, details with respect to the computational complexity are covered in this section.

Typically there are two sets of computations associated with every classification technique – computations needed for training and computations required for labelling new data. The first part is in essence a one-time offline procedure that is carried out before the deployment of the classifier in practice. On the other hand, data labelling is the actual computation procedure that takes place in the real-life operation. Therefore in the computational complexity analysis we have considered only this part. The computational complexity for each of the classifiers is expressed in terms of their required arithmetic operations as they are representative of energy consumption of each classifier. Since several implementations of the same arithmetic function and the classifier architecture are possible, to put them on a uniform platform we consider flat unfolded architecture without any resource sharing or parallelism. Thereby computational complexity can be described as the fundamental numbers of arithmetic operations required for each of them. Care should be taken for k-NN, though, as arithmetic operations for determining the k nearest neighbours from the full set of training samples are also considered, apart from those involved in computing the distance. Also, for simplicity, complexity of subtraction operation is considered equal to that of an addition. The arithmetic complexities of different classifiers thus derived are shown in Table 5.1. Here N, M, S indicate the dimension of feature vector, the number of SVs in SVM and the number of training samples respectively.

From Table 5.1 it is evident that with a given number of feature vectors, LDA exhibits least computational complexity whereas the same in SVM and k-NN strongly depends on the number of SV and training samples used respectively. In addition, with the

Table 5.1: Total number of arithmetic operations involved in labelling a new sample for the five classifiers.

	Add (+)	**Mul (×)**	**Sqr ($()^2$)**	**Sqrt ($\surd$)**
LDA	N	N	0	0
QDA	N^2+N	N^2+N	N	0
SVM$_L$	NM	$(N+2)M$	0	0
SVM$_Q$	$(N+1)M$	$(N+2)M$	M	0
k-NN	$2S(N+1)-6$	0	SN	S

increase of feature vectors, LDA effectively shows linear increase in computational complexity; on the other hand, all the others approximately exhibit quadratic increase of the same (considering M and S comparable to N and squaring operation equivalent to multiplication). It is to be noted that as an arithmetic block, a multiplier (and a divider) is several times more energy consuming than an adder. Therefore, due to less number of multipliers required to realise LDA, it is expected that LDA will consume much less energy than the other classifiers.

In order to create an unified metric describing overall computational complexity for each of the classifiers, we used 2-input NAND gate complexity (NG). Considering unfolded architecture, no resource sharing for each of the arithmetic computational modules and b-bit wordlength implementation, the numbers of transistors required for each of them can be given as $T_+ = 24b$, $T_\times$ and $T_{()^2} = 30b^2 - 36b$, $T_{\surd} = 18(\frac{b}{2}+1)(\frac{b}{2}+3)$ [225], where T_* denotes the transistor count for the arithmetic operation (*). Since a 2-input NAND gate requires 4 transistors, these numbers could be transformed into NAND gate equivalent as $G_+ = 6b$, $G_\times$ and $G_{()^2} = \frac{15}{2}b^2 - 9b$ and $G_{\surd} = \frac{9}{2}(\frac{b}{2}+1)(\frac{b}{2}+3)$. In our study, a word length of $b = 16$ bit is considered for the sake of demonstration.

5.2 Heartbeat Classification

To exemplify spectral energy-based classification with and without more features, two scenarios are considered: single heartbeat classification and multiple heartbeat classification. *Single heartbeat classification* focuses on one set of heartbeats (i.e. one simultaneously captured beat per participating lead, as spectral energy were derived on lead basis and depending on the lead scenarios it could be more than one heartbeat). It is similar to what we did in Section 4.3. Building upon this concept, *multiple heartbeat classification* is also examined. It refers to *multiple* single heartbeats, where multiple simultaneously captured and consecutive beats per participating lead are present.

Single heartbeat classification is necessary in our study because it lay down the fundamental judgement for assessing spectral energy as a feature and classification performance. To take the justification for experimental results even further, multiple heartbeat classification is required. That is also because, from signal morphological point of view, relying only on a single heartbeat to judge ECG classification for good or bad may not be accurate or practical. Occasionally in clinical practice, an abnormal heartbeat may occur in isolation (like ectopic beat) along with some other beats, but in fact the patient is diagnosed as normal. By just looking at one heartbeat may lead to wrong decision. That is also why clinical doctor tend to make a decision on the basis of multiple heartbeats when examining the ECG paper. Analogically, in our case, if this isolated abnormal heartbeat was chosen for analysis, the overall diagnosis will deem to be wrong. As a result, it is always advisable to consider multiple heartbeats per lead basis for classification for the sake of consistency check.

5.3 Spectral Energy-based Classification

In this section, spectral energy-based classification without augment of any features is discussed. As for single heartbeat classification, it will be primarily based on Set 4 in Section 4.3.5 in terms of experimental procedure. But more importantly, an elaborate experimental discussion on the trade-off between classification performance and computational complexity will be made, especially for SVM. In addition, multiple heartbeat classification is also carried out as an extended study to examine the effectiveness of the outcome of single heartbeat classification.

5.3.1 Single Heartbeat Classification

Experimental procedure and analysis of trade-off between accuracy and computational complexity will be discussed, following which multiple heartbeat classification will be given.

5.3.1.1 Experimental Procedure

As we have already shown in Section 4.3, DWT based spectral energy of the ECG beat components (Set 4) has been proved to be an effective approach of extracting the distinctively characteristic features for ECG classification. On this basis, this approach has been taken as the primary means of feature construction for our spectral energy-based classification for both single-beat and multiple-beat scenarios. The way to construct this set of features should resort to the entire procedure discussed in Section 4.3.5, which eventually leads to the set of features for each lead scenario listed in Table 4.6. In the

following, we will perform a more elaborate classification performance evaluation, with respect to the previous experiments in Section 4.3.5, with a focus on computational complexity of making prediction on new sample.

5.3.1.2 Trade-off between Accuracy and Computational Complexity

Despite we have already derived the classification results for Set 4 in Table 4.7, this time a more elaborate cross validation with 10 runs of 10-fold is executed to obtain consistent classification performance of the classifier. On the other hand, calculation of computational complexity is also needed. However, cares should be taken here. As already shown in Section 5.1, in theory LDA is expected to be computationally least complex. Conversely, SVM is fairly complex, though claimed to be of high classification performance [192]. However, parametrical turning of SVM may be possible to reduce the complexity at the expense of accuracy in presence of the current database. Investigation into such may unveil optimised SVM. In that case fair comparison between classifiers can be made.

As a separate investigation, SVM will be discussed following LDA, QDA and k-NN. So first of all, classification performances and computational complexities of the three are derived and listed in Table 5.2. As we can see, on average specificity seems to be higher than sensitivity to varying extent depending on the lead scenario and classifier, except QDA. More importantly, the overall assessment of classification – accuracy shows an increasing trend along with the lead scenario. Effectively this attributes to the growing dimension of feature space under lead scenarios, by adding useful features from extra leads. Same thing happens to computational complexity.

Now, apart from the similar reasons, classification performance and computational complexity of SVM are also affected by the number of SVs needed. To increase/decrease the number, the regularisation parameter C_s may be tuned [185] so that direct influence to both the performance and complexity can be made. Therefore, we attempt to explore the trade-off between accuracy and the number of SVs. Initially, C_s was set to 1, as it has been used to find out the best lead (and feature) combination in Section 4.3.5. Now, C_s is subjected to grid search ($C_s = 2^i$, $i = -15, -14 \ldots 14, 15$) during the training phase. In particular, C_s^{min} that achieves the minimum number of SVs while producing acceptable performance will be preserved for every lead scenario. To distinguish the two from experimental outcome perspective, two cases have been set up:

- *Case 1*: it is a scenario where maximal accuracy without concerning about the number of SVs deployed. Here C_s is set to 1 as we did initially.
- *Case 2*: it is a scenario where minimum number of SVs and the associated accuracy are obtained. Here C_s follows grid search and C_s^{min} is found and preserved.

Table 5.2: Classification performance and associated computational complexity for LDA, QDA and k-NN. Best result of the metrics (in column) among lead scenarios for the same classifier is boldfaced.

	Lead Scen.	**Spe**(%)	**Sen**(%)	**Acc**(%)	**NG**(log_{10})
LDA	**1**	84.81	64.04	74.42	3.7494
	2	81.54	**86.92**	84.23	3.8743
	3	93.27	83.65	88.46	4.1754
	4	**93.46**	84.81	89.13	**4.2265**
	5	92.88	85.96	**89.42**	4.1754
QDA	**1**	63.08	85.96	74.52	4.1698
	2	68.27	**89.23**	78.75	4.9507
	3	**83.08**	82.50	82.79	5.0691
	4	81.92	85.00	83.46	**5.2659**
	5	80.96	87.69	**84.33**	**5.2659**
k-NN	**1**	78.85	80.38	79.62	5.3833
	2	91.54	84.42	87.98	5.6281
	3	90.19	82.12	86.15	6.1270
	4	90.00	84.23	87.12	6.1826
	5	**94.62**	**85.96**	**90.29**	**6.2762**

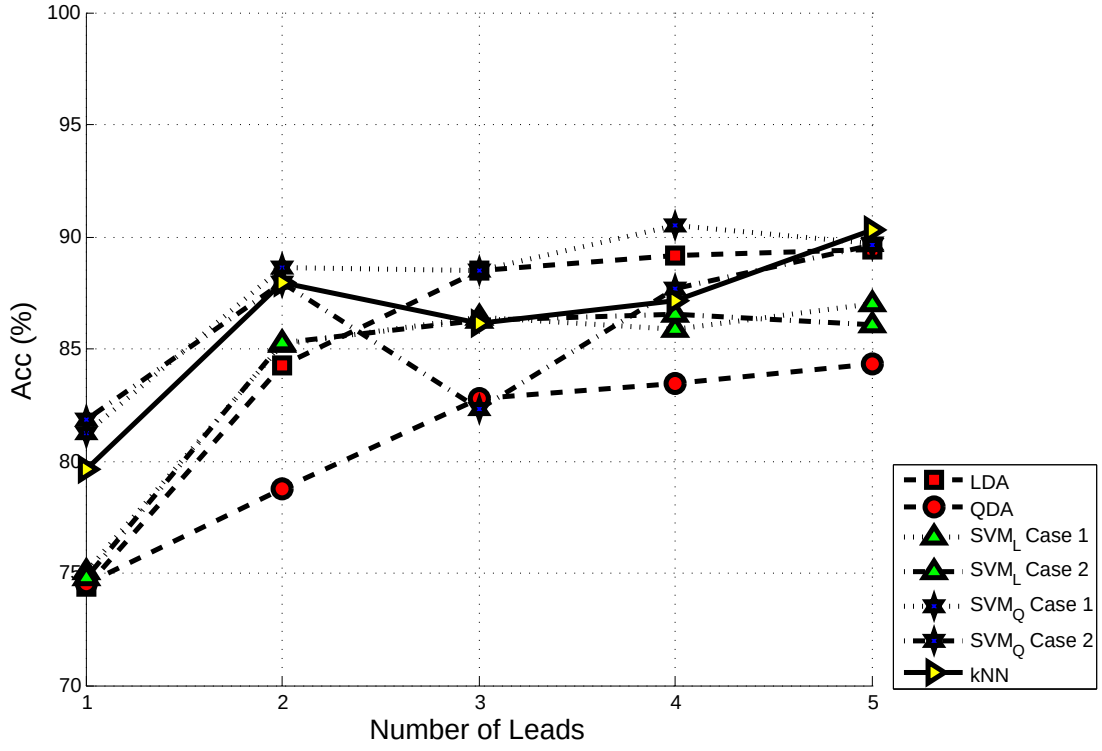
Table 5.3: Classification performance, number of SVs, C_s and associated computational complexity for SVM. Best result of the metrics (in column) among lead scenarios for the same classifier is boldfaced.

	Lead Scen.	**Spe**(%)	**Sen**(%)	**Acc**(%)	**# of SVs**	C_s	**NG**(log_{10})
SVM$_L$	**1**	61.15	**89.04**	75.10	86	1	**5.8032**
(Case 1)	**2**	81.73	88.65	85.19	72	1	5.8857
	3	**89.81**	83.08	86.44	72	1	6.0115
	4	85.19	86.54	85.87	64	1	6.0555
	5	88.08	85.96	**87.02**	**60**	1	6.1056
SVM$_L$	**1**	60.96	**88.65**	74.81	86	4	**5.8032**
(Case 2)	**2**	82.31	88.27	85.29	68	2048	5.8598
	3	**90.58**	81.92	86.25	68	512	5.9856
	4	86.15	86.92	**86.54**	58	64	6.0105
	5	87.88	84.23	86.06	**50**	32	6.0736
SVM$_Q$	**1**	76.15	86.35	81.25	83	1	**5.8700**
(Case 1)	**2**	90.77	86.54	88.65	59	1	5.8850
	3	92.50	84.42	88.46	50	1	5.9690
	4	**94.04**	86.92	**90.48**	44	1	5.8830
	5	91.92	**87.50**	89.71	**40**	1	5.9136
SVM$_Q$	**1**	77.50	86.15	81.83	81	128	5.8752
(Case 2)	**2**	90.77	85.19	87.98	40	64	5.7473
	3	85.00	79.62	82.31	31	16	5.6686
	4	89.04	86.35	87.69	25	64	5.6679
	5	**92.31**	**86.92**	**89.62**	**24**	1024	**5.6341**

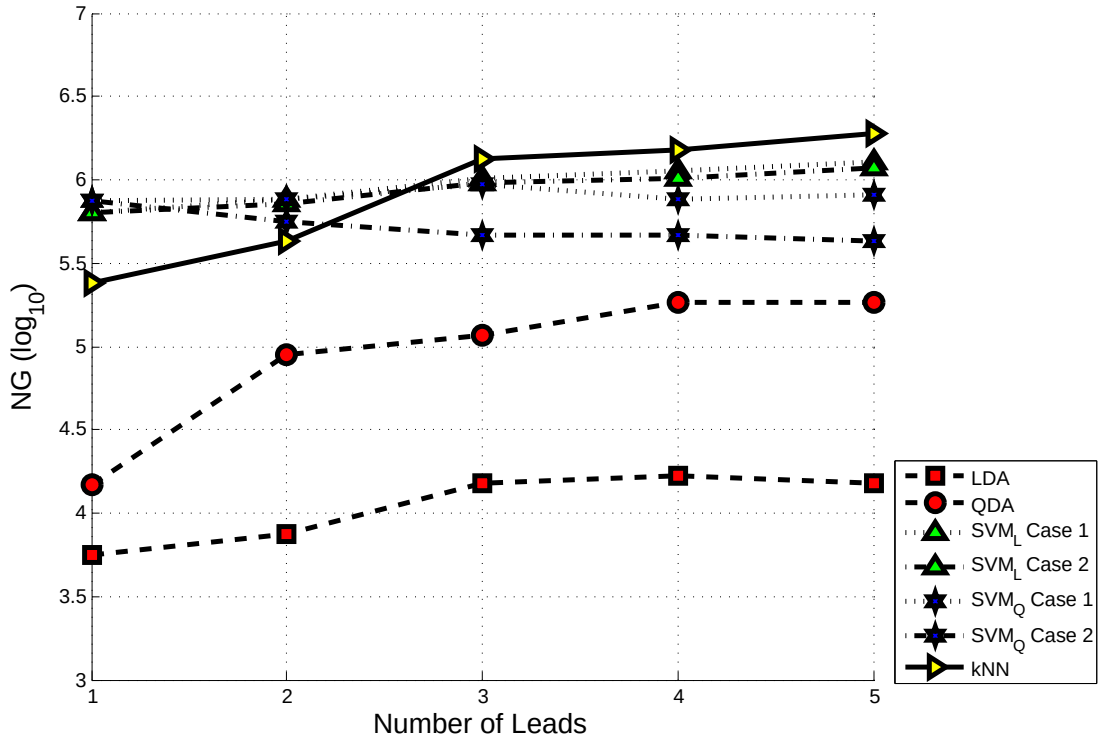
Table 5.3 shows the classification performance and computational complexity for both cases of SVM. Positive and negative gap between Case 1 and Case 2 in terms of specificity and sensitivity can be observed for both SVM. Accuracy affected by these two metrics grows along with the lead scenario, which is expected. Also, in general it exhibits slight degradation to the degree of mostly 1% from Case 1 to Case 2, which is also expected. More importantly, by tuning the C_s the optimal number of SVs is achieved, allowing reduction to be made upon computational complexity. Although not much reduction ($<$ 0.1) of computational complexity is observed in SVM_L due to small decrease in number of SVs, relatively huge reduction ($>$ 0.3) can be observed in SVM_Q particularly in 3 to 5 lead scenarios. That effectively tells that trade-off via parametrical tuning allows us to balance accuracy for computational complexity without impairing much accuracy. However, the effectiveness of trade-off still depends on the setting like features involved, kernel used in SVM, etc.

After obtaining the two important metrics – accuracy and computational complexity for the five classifiers, we need to combine the views of the two so that comparison can be made between classifiers. To illustrate the results for easy understanding, Figure 5.1 depicts the trends of accuracy as well as the computational complexity with respect to lead scenario for all classifiers. Focusing on the accuracy (Figure 5.1(a)), clearly overall trend is growing thanks to the increasing number of features involved. Specially, QDA is seen to perform to the lowest level with regards to other classifiers; whereas highest can be seen in SVM_Q (Case 1), followed by k-NN in general. Interestingly enough, though with some gaps in 1 and 2 lead scenarios, LDA seems to perform comparably well as SVM_Q does in 3 to 5 lead scenarios. This observation leads us to think of the computational complexity they might take. In regards to complexity (Figure 5.1(b)), though SVM_Q (Case 1) has shown great classification performance, its complexity is of around two order of magnitude higher than LDA. Even with its Case 2, complexity is nonnegligible to LDA. Same applies to k-NN, and its minor advantage of accuracy over LDA may not be appreciated by its computational complexity. Without any advantages of accuracy or computational complexity, the other two classifiers (SVM_L and QDA) are not comparable to LDA at all. Notice that complexity of LDA and SVMs decreases as the number of lead increases. This is because (1) for LDA the number of features in 5 lead scenario is actually smaller than 4 lead scenario; (2) for SVMs the number of support vectors in fact decreases along the increase of leads.

Overall, the results for both accuracy and computational complexity indicate that complexity taken by LDA is significantly lower than those of others; meanwhile the accuracy it achieves is either the highest or within small margin of the best, except 7% less accurate than SVM_Q in 1 lead scenario. Attempts to further justify the single heartbeat classification will be made in the following section.



(a)



(b)

Figure 5.1: (a) Raw classification accuracy (Acc) versus number of leads for all classifiers; (b) Computational complexity (log_{10} form of NG) versus lead scenario for all classifiers.

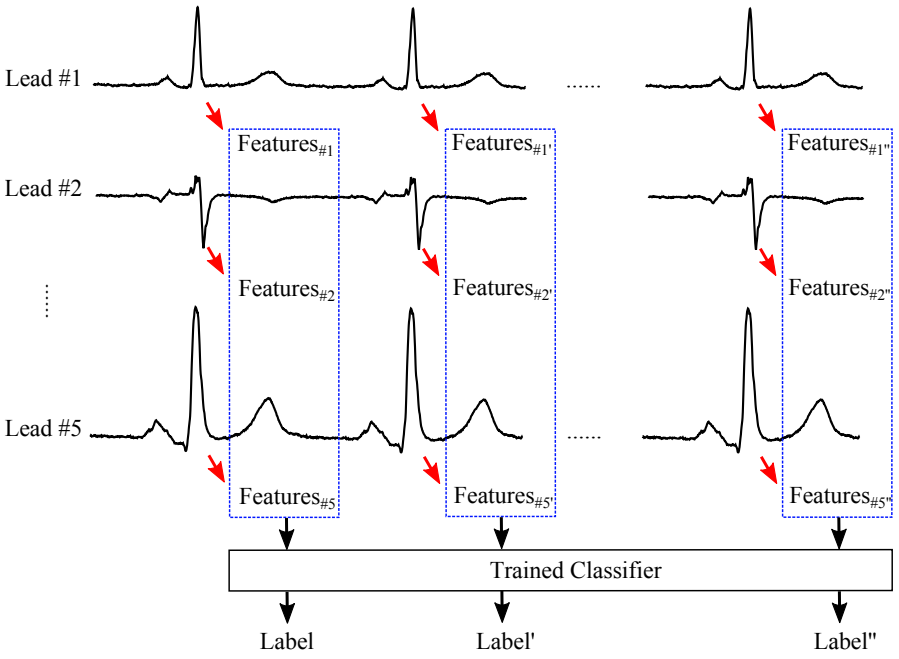


Figure 5.2: An illustrative demonstration of multiple heartbeat classification.

5.3.2 Multiple Heartbeat Classification

As an extended study of single heartbeat classification, multiple heartbeat classification is carried out in this section. In principle, we follow the same flow of feature construction in Section 5.3.1.1. However, for multiple heartbeat case, we also need to segment out each heartbeat within the multiple heartbeats before we apply the feature construction steps on each of them. In addition to this, since the lengths of the signals in our two ECG databases (Section 4.3.1) vary, the total number of heartbeats for each lead and for each patient is ultimately set to 7. This is due to the fact that the minimum number of heartbeats available amongst the 104 records was found to be 7. Next, an ECG segmentation algorithm is applied to separate out each heartbeat amongst these 7 heartbeats, and then TDMG (Section 3.2) is invoked to obtain the boundaries of P, QRS and T waves within each of the heartbeats. Notice that those heartbeats that occur to be problematic with wave boundary detection algorithm are ignored and the next applicable beat is included instead. Doing so does not affect the classification performance in that all the heartbeats equally exist and they should be treated equally. Subsequently, with the outcome of TDMG, spectral energy of interest can be derived for multiple-beat scenario.

To perform prediction on multiple-beat scenario, classification model has to be trained beforehand. Here we directly utilise the trained models from single-beat classification, where the regular and most representative heartbeat of each patient was selected and used to train the classifiers. Clearly, depending on different lead scenarios, classifier must be chosen differently. As we have already shown that the best feature space

each classifier could achieve the highest total accuracy with is their respective 5 lead scenario, it is fairly logical for us to focus on this particular lead scenario and expect to have the best accuracy in multiple heartbeat analysis. Therefore, we opt for the trained classifiers of 5 lead scenario. Accordingly, multiple heartbeat classification is carried out with these classifiers, and thereby classifying the 7 heartbeats of each patient into normal or abnormal classes. To illustrate, Figure 5.2 presents the process of multiple heartbeat classification, in which features (in this case spectral energy) are derived from each beat under 5 lead scenario and subject to the trained classifier to produce a label. This loops until it reaches the seventh heartbeats and thereby seven labels are obtained.

To finally make a decision on whether the patient is normal or abnormal on the basis of these 7 heartbeats, one simple decision-making scheme is applied in our analysis: by counting the number of normal and abnormal labels, the one that has more normal labels is classified as normal while the one that has more abnormal labels is classified as abnormal. However, one may argue that if a patient is abnormal because of 1, 2 or 3 beats, then the system would not be able to detect the abnormality. Recall that what we define as abnormal patient (see Section 1.1) is of low chance in having just 1, 2 or 3 abnormal heartbeats. For acute type, they tend to have at least more than 3 abnormal beats (e.g. acute MI); for most of the chronic type, it also tends to have more than 3 abnormal beats (e.g. patient with scar). That means for abnormal patient in our study, it actually makes sense to make the decision based on 7 heartbeats, because they tend to exhibit more than 3 abnormal heartbetas in sequence.

To evaluate the performance of each classifier, we used sensitivity, specificity and raw classification accuracy. Simulation has been done and the associated results can be found in Table 5.4.

Table 5.4: Simulation results of multiple heartbeat classification for all classifiers.

	LDA	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
Spe (%)	84.31	84.31	43.14	66.67	72.55
Sen (%)	86.96	84.78	91.30	89.13	84.78
Acc (%)	85.57	84.54	65.98	77.32	78.35

In terms of specificity, it can be seen that LDA performs at 84.31% as QDA does. Surprisingly, the other three classifiers achieve less to different degrees, where 43.14% for SVM$_L$, 66.67% for SVM$_Q$ and 72.55% for k-NN. In terms of sensitivity, SVM$_L$ and SVM$_Q$ turn out to only surpass their opponents between 3% to 7%. That means LDA and QDA perform fairly good at both metrics, indicating that in most cases they are able to predict normal and abnormal heartbeats correctly for individual patients. On the other hand, both SVMs performance are inferior in predicting normal heartbeats, but conversely good at abnormal prediction. k-NN, as a very computationally complex classifier amongst the rest particularly in higher lead scenario, do not perform well as

expected. Overall, because of the better performance in specificity, LDA and QDA achieve relatively better accuracy than the other three classifiers, at 85.57% and 84.54% respectively. That means less-complex classifiers like LDA and QDA seem to perform fairly well in multiple heartbeat classification, even better than higher-complex classifiers like SVM and k-NN. In addition, note that all the classifiers show reduced accuracy compared to single-beat case (Table 5.2 and Table 5.3). This may stem from two reasons, that features we have drawn from a single beat for classifier trainings may depend too heavily on this specific beat to generalise to various beats; and that spectral energy as features may not be sufficient to achieve better, or even maintain, the performance.

5.3.3 Discussion

Having obtained the results of spectral energy-based single heartbeat and multiple heartbeats classification, it is not difficult to see LDA has shown promising performance in classification as well as computational complexity. With comparably high classification performance and significantly low complexity, so far LDA in 5 lead scenario has rendered itself the best candidate for normal and abnormal heartbeat classification amongst the rest. However, improvement on classification performance may be possible when more features apart from spectral energy join in, particularly for multiple heartbeats classification. In theory, deploying more relevant features helps increase generalisation. That also means, improvement is expected on all 5 classifiers. It might so happen that, with more features, other classifiers happen to outperform LDA while keeping computational complexity low. In this case, LDA might not be a good option for hardware implementation. Therefore, decision on which classifier should be selected for hardware implementation may not be possible until more solid investigation is done, and it leads us to the next section.

5.4 Spectral Energy-based Classification Augmented with More Features

Having covered spectral energy-based classification without augment of any features, we opt to examine spectral energy-based classification with augment of more features in this section. As for single heartbeat classification, we will directly treat the spectral energy feature space (Table 4.6) as the basis, and upon which more features will be added accordingly through some feature selection processings. Detailed investigation into how these extra features are generated and selected will be covered. Classification performance and the associated computational complexity will also be covered. In addition, analysis of multiple heartbeat classification is done as an extended study to examine the effectiveness of the outcome of single heartbeat classification.

5.4.1 Potential Features for Classification Enhancement

In the following, we will attempt to extract more features from ECG heartbeat, mainly in time domain as well as frequency domain including DFT and DWT. Reasons of introducing more features from these three specific domains are the following. Firstly, it is logical to bring in time domain features, in that they closely relate to clinical knowledge and are expected to facilitate the classification performance quite well. Secondly, we also opt for DFT domain. But it may be argued that, there is no point in deploying DFT as it takes huge amount of computational complexity and will eventually consume a lot of power. This could make related features impractical to compute in mobile environment. However, still we would like to include it because (1) DFT-based approach has been regularly reported [135, 136, 210, 226, 227] in parallel with wavelet-based approaches in ECG classification for the past two decades. Clearly it still demonstrated its usefulness in classification; (2) comparison between DFT- and DWT-based approaches can be directly made in this specific study. Thirdly, as discussed in Section 2.2.2, natural effectiveness of DWT in non-stationary signal processing is undoubted. Features built upon DWT are expected to take the advantage. In the following, we will explain what features from these three domains are extracted.

In regards to time domain features, we look for duration of P, PR, QRS, QT and T waves, and also peak amplitude of P, Q, R, S, T waves. This is because they are directly available to us from feature detection algorithm (Chapter 3). Note that to obtain peak amplitude, one must subtract the absolute amplitude with respect to baseline level so that relative amplitude can be obtained. Here, we simply treat the onset of P, QRS, T as baseline for corresponding waves, respectively. That is because in reality they are expected to lie on baseline as TP segment normally does (Figure 2.5).

As for DFT, Figure 4.1 serves as the guide for extracting the features. According to the figure, low-frequencies like P and T waves dominate within the 5 Hz frequency bins, whereas QRS complex mainly spread through the frequency bins from 5 Hz up until 40 Hz. As a result, it is sensible to derive features for P and T waves as well as QRS complex within their own frequency bins. So, as for DFT-based features, deploying periodogram upon an entire heartbeat signal leads us to onesided Power Spectrum Density (PSD). From there, we follow suggestions in [228, 226, 140] and opt to derive (1) the frequencies that achieve the maximum power for PT waves and QRS complex respectively; (2) the mean and standard deviation of power for both within their own bins, as well as the mean and standard deviation of power for the entire spectrum (0.5 to 40Hz); (3) the half point of energy (or power), i.e. the frequency that divides up the spectrum into two parts of the same area; and (4) spectral entropy of DFT as the distribution of energy described by probability, following Equation 5.1

$$H_{DFT} = -\sum_i p_i \log_2 p_i, \text{ where } p_i = \frac{|F_i^2|}{\sum_k |F_k|^2}, \ i,k = 1,\dots N \tag{5.1}$$

where F_i denotes the DFT coefficient. Note that a so-called ordered activity (e.g. a sinusoidal signal) manifests as a narrow peak in frequency domain. This in fact corresponds to low entropy. However, disordered activity signal like noise would lead to higher entropy as a wide band response exists accordingly.

On the other hand, DWT-based features are also of our interest. As suggested by [209], mean and standard deviation of detail coefficient sequence at each decomposition level are derived for our purpose. In addition, DWT entropy has shown effectiveness and attracted huge attention in brain signal analysis [208]. Now we extend its use to our study. With Equation 2.11 and 4.6, we can derive total wavelet energy (Equation 5.2) and relative wavelet energy (Equation 5.3).

$$E_{TW} = \sum_{m=1}^{N} E_{cD}^j + E_{cA}^N, \text{ where } E_{cA}^N = \sum_{n=1}^{N} |cA[n]|^2 \tag{5.2}$$

$$E_{RW}^j = \frac{E_W^j}{E_{TW}} \tag{5.3}$$

where N is the maximum decomposition level. Note that wavelet energy of E_{cA}^N is also needed in calculation of total wavelet energy as the energy of the approximate DWT coefficient sequence at the highest level is required to construct the total energy of the signal. With both, level-based DWT entropy (Equation 5.4) and the total DWT entropy (Equation 5.5) can then be derived as

$$H_{LW}^j = -E_{RW}^j \cdot \log_2(E_{RW}^j) \tag{5.4}$$

$$H_{TW} = \sum_{j=1}^{N} H_{LW}^j \tag{5.5}$$

Both kinds of entropy reflect the degree of order and disorder of the signal. More specifically, Equation 5.4 and Equation 5.5 act as a measure of the information of the DWT coefficient distribution on certain-level scale and entire scale respectively. Due to the nature of DWT, noises embedded within the signal can also be easily eliminated. As a result for level-based DWT entropy, we focus on level 2 to 5 (i.e. 4 levels in total).

Table 5.5 summaries the features discussed above, which are categorised into time domain (morphological) as well as DFT domain and DWT domain (statistical). Bear in mind that features in Table 1 are not limited to one heartbeat in one lead. Instead, synchronised heartbeats in 12 leads are all subjected to the extraction of those features. In other words, for each patient, each type of feature is extracted from 12 leads. That

Table 5.5: Features distributed by categories: Time domain, DFT domain and DWT domain.

	Features	#
Time Domain	• Duration of P, PR, QRS, T, QT waves; • Peak ammplitude of P, Q, R, S, T waves.	120
DFT Domain	• MaxPowFreq of PT and QRS; • Mean and StdDev of PT, QRS and the entire spectrum; • Half point of energy; • DFT entropy.	120
DWT Domain	• Mean and Std at DWT level 2 to 5, respectively; • Total wavelet entropy; • Level-based wavelet entropy at level 2 to 5 respectively.	156
Total		396

means, in total there are 396 features constructed for latter feature selection processing.

5.4.2 Feature Selection Algorithms

Having obtained the features in last section, decision should be made on which feature selection algorithms will be used to serve our purpose. To do so, following the unifying platform proposed in [157] allows us to find out a proper strategy of deploying necessary feature selection algorithms. From there, time limit and the purpose of feature selection effectively direct us to filter method, not wrapper nor hybrid method. The reason is that, though it might be beneficial to follow the latter two as they are better suited to learning algorithms and hence better performance, it may take huge amount of time to get the training done (in particular, SVM optimisation) in the cross validation phase, making it fairly impractical. On this basis, we opt for filter method.

In addition, a three-dimensional framework, including search strategies, evaluation criteria and data mining tasks, for categorisation of feature selection algorithms is also proposed in [157]. Specifically on each dimension, search strategies are divided into complete, sequential and random search; evaluation criteria for filter are divided into distance, information, dependency and consistency; and data mining tasks are divided into classification and clustering. As for our study, we are keen to deploy feature selection algorithms that cover different search strategies and different distance measurement schemes tailored to classification. Also, the prospective algorithms are expected to cover uni- and multi-variate situations, two different forms of output (feature weighting and feature set), and the capability to handling redundancy. Selecting proper feature selection algorithms can therefore be achieved in this way, with a hope of finding relevant features in a time-saving and comprehensive way. Among the big pool of feature selections algorithms [157, 229], this effectively narrows down the choices and reach to the

following algorithms: ReliefF, Information Gain (InfoGain), Correlation-based Feature Selection (CFS) and Fast Correlation-Based Filter (FCBF). More details about these four algorithms will be covered in the next section.

5.4.2.1 Preliminaries of the Four Feature Selection Algorithms

In the following, a brief of the basics of the four feature selection algorithms is given. All these algorithms are publicly available in [158] and [230].

ReliefF Originally *Relief* was proposed in [231] and later its multiclass extension ReliefF [232] was introduced. It is a technique that estimate features according to how well their values distinguish among samples that are near each other. The algorithm effectively searches for two nearest neighbours of a given sample: one from the same class and the other from different class, and according to the evaluation criterion

$$RF = \frac{1}{2}\sum_{n=1}^{T} d(\mathbf{x}_{n,i} - \mathbf{x}_{ND(\mathbf{x}_n),i}) - d(\mathbf{x}_{n,i} - \mathbf{x}_{NS(\mathbf{x}_n),i}) \tag{5.6}$$

where $\mathbf{x}_{n,i}$ denotes the dimension i (feature i) of sample $\mathbf{x}_n$, and $\mathbf{x}_{ND(\mathbf{x}_n),i}$ and $\mathbf{x}_{NS(\mathbf{x}_n),i}$ denote the dimension i of the nearest samples to $\mathbf{x}_n$ but with different and same class label, respectively. $d(.)$ is the distance measurement in ReliefF. Depending on such evaluation criterion, the weight of the features for ranking will be assigned and therefore feature selection is achieved. This algorithm chooses features that devote mostly to the separation of the samples from different classes. As ReliefF covers two-class situation, which is the case in our study, we choose to use ReliefF straightaway.

Infomation Gain *Information gain* (InfoGain) [233] is a measure of dependence between the feature (or variable) and the class label. It is one of the most popular feature selection techniques due to its easy computation and interpretation. Equation 5.7 shows how to calculate information gain[2]

$$IG(X|Y) = H(X) - H(X|Y) \tag{5.7}$$

where

$$H(X) = -\sum_{j} p(x_j)\log_2(p(x_j)) \tag{5.8}$$

[2]To generalise the equation without tailoring to any specific symbols, we use X and Y here to denote two different generic variables X and Y. That means X can be any feature, and Y can be any feature or any class label, in a mathematical sense. This is mainly done to avoid confusion when deployed in CFS and FCBF.

$$H(X|Y) = -\sum_{k} p(y_k) \sum_{j} p(x_j|y_k) \log_2(p(x_j|y_k)) \tag{5.9}$$

$H(X)$ and $H(X|Y)$ are the *entropy* of X and entropy of X after observing Y, respectively. Note that X can be substituted as a feature x of class j and Y can be substituted as class label y with index k. According to the equations, a feature with high IG value is relevant, or irrelevant otherwise. In other words, if $IG(X|Y) > IG(Z|Y)$, that means feature Y is considered to be more correlated to feature X than to feature Z [234]. Direct usage of Information Gain is possible to provide a ranked list of features; also another typical use of Information Gain is in constructing decision tree.

CFS By using a correlation based heuristic approach, Correlation-based Feature Selection (CFS) [235] is able to evaluate the worth of features. Such heuristic takes into account the usefulness of individual features for predicting the class label along with the level of intercorrelation among themselves. To formalise the heuristic, we have

$$\text{Merit}_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{5.10}$$

where Merit_S is the heuristic merit of a feature subset S containing k features; $\overline{r_{cf}}$ is the mean feature-class correlation ($f \in S$); and $\overline{r_{ff}}$ is the average feature-feature inter-correlation. Numerator of Equation 5.10 indicates how this group of features are good at predicting the class; on the other hand, denominator tells the level of redundancy between the features. Moreover, correlations of features are effectively estimated based on the information theory, thus the degree of association between nominal features. Here, Information Gain (Equation 5.7) is deployed to achieve this. Furthermore, Best First search [236] as the method of searching the feature subset space is used. Basically CFS starts from the empty set of features and uses a forward best first search. It only stops when five consecutive fully expanded non-improving subsets are found. Advantage of CFS is that it works well on smaller data sets and avoids redundancy.

FCBF Fast Correlation-Based Filter (FCBF) is a filter-based feature selection algorithm that estimates the feature-class and feature-feature correlation [234]. Rather than adopting classical linear correlation approach as correlation measure, FCBF opts for information gain based on entropy. It is similar to CFS in this sense, but with a different evaluation criterion, which is symmetrical uncertainty.

$$SU(X,Y) = 2\,[\frac{IG(X|Y)}{H(X) + H(Y)}] \tag{5.11}$$

Table 5.6: Main characteristics of the four selected feature selection algorithms.

	Search Strategy	**Evaluation Criterion**	**Output Type**	**Variate**	**Handle Redundancy?**
ReliefF	Sequential	Distance	Feature weighting	Uni	✖
InfoGain	Complete	Information	Feature weighting	Uni	✖
CFS	Sequential	Dependence	Feature subset	Multi	✔
FCBF	Sequential	Information	Feature subset	Multi	✔

Symmetrical uncertainty is aimed to compensate for information gains bias, as otherwise information gain would be biased to features with more values. Next, using symmetrical uncertainty as the goodness measure, a sophisticated procedure that decides whether a feature is relevant to the class and whether such a relevant feature is redundant or not when considering it with other relevant features is carried out. Overall, this method is capable of handling feature redundancy very effectively, and also lends itself well to time complexity compared to others, e.g. ReliefF.

Table 5.6 concludes the main features of our four selected algorithms. All of them are supervised. Note that, in regards to the difference between univariate and multivariate, it actually refers to whether an individual feature (univariate) or a subset of features (multivariate) should be added or deleted at each round of feature selection, i.e. during subset evaluation (Section 2.3.3.2) [237]. For univariate algorithms like ReliefF and InfoGain, they evaluate features individually and therefore cannot handle feature redundancy. In addition, for output type, the difference between the two is about the order among the selected features. In feature weighting, ranked list of the features exists, whereas feature subset does not. That means one can easily remove the least relevant features in the first case, but not expected to do so in the second case.

5.4.3 Single Heartbeat Classification

Similarly to Section 5.3.1, experimental procedure as well as the associated results will be discussed, but with focus on augment of more features. After that, multiple heartbeat classification will be covered.

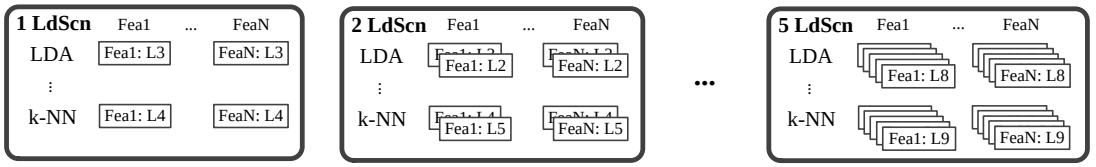


Figure 5.3: Choosing features according to spectral energy based lead scenarios from certain feature group.

5.4.3.1 Experimental Procedure

Having introduced potential features and feature selection algorithms, we need to follow a systematic experimental procedure to achieve our goal. In the following, the entire process is described in order.

1. *Standardisation*: every feature is standardised as preprocessing step according to Equation 2.12. Therefore all features comply with zero mean and unit variance.

2. *Feature space construction for potential features*: before we run any feature selection algorithms, care should be taken in constructing feature space. First of all, features should be collected into different groups considering the fundamental computation methods required (hence computational complexity required), so that features derived from one method would not be mixed with others. Here, Table 5.5 offers us good divisions of the features. Besides, since we have particular interest in wave duration (as shown in Section 4.4), thereby action is made to have duration and peak as two extra groups. Hence, in total we have five groups of potential features: Duration, Peak, Time domain, DFT domain and DWT domain.

   Secondly, as we know, for each feature we have 12 leads of it. But instead of all 12 leads, we opt to pick up those that share the same lead(s) with spectral energy based lead scenario (Table 4.6). Doing so *constraints* ourselves to features under these fixed lead scenarios, and will in turn facilitate the feature selection phase later. Otherwise we might end up with useful features after feature selection but from leads that are different to existing lead scenarios, which eventually would be of no use. To illustrate the idea, Figure 5.3 takes one feature group as an example and depicts how we pick up features accordingly. In a feature group, there are feature1 to featureN. Under each lead scenario, features from leads (shown in Table 4.6) for certain classifier (e.g. LDA) are preserved. For instance, under 1 lead scenario feature1 in lead 3 is preserved for LDA. So does featureN in lead 3. Having lead 3 here is because it is the only lead participated in 1 lead scenario for LDA in Table 4.6. Similarly, lead 4 is the only lead in 1 lead scenario for k-NN, hence feature1 to featureN in lead 4 are preserved accordingly. Expanding the idea further to 5 lead scenario, it is not difficult to see only corresponding features

are preserved when their leads are shown in Table 4.6. Doing so allows us to link and facilitate the spectral energy based classification by only considering features from the same leads as spectral energy does. Note that all five feature groups are subjected to the same strategy shown in Figure 5.3.

3. *Executing feature selection*: once features are properly arranged, feature selection algorithms can be executed to rank/select the features, as shown in Figure 5.4. To explain, let us take 1 lead scenario from Figure 5.3 as an example. As our target is to facilitate spectral energy under five classifiers, we treat features under different classifiers as individual units. After that, features in units are fed into our four feature selection algorithms, thereby ranked features (i.e. features with different weights) and optimal feature set can be achieved under corresponding feature selection algorithms. The above process iteratively applies to all lead scenarios and further all feature groups.

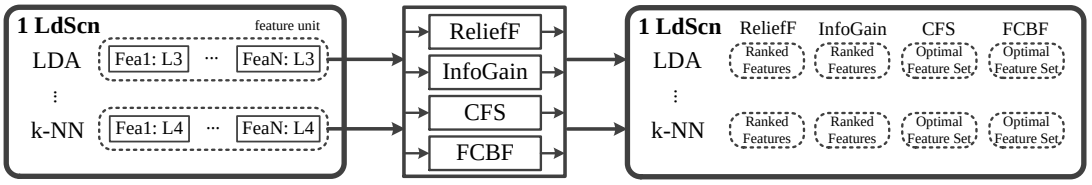


Figure 5.4: Executing feature selection on potential features. (Here 1 lead scenario is taken as an example.)

4. *Coupling spectral energy based feature space with new features*: having obtained the outcome of feature selection algorithms, it is time to couple spectral energy based feature space with our new features. Since we have covered two different output types of feature selection, strategies of coupling the features to spectral energy must be taken differently. As for ReliefF and InfoGain, features in the same unit are ranked with respect to their weights. Here higher weight represents higher relevance to the class. Thus, feature with the highest weight is joined with features of spectral energy under the same classifier and lead scenario for classification evaluation, upon which the second highest one will then join and new evaluation will be carried out. It keeps running until all ranked features are taken into account. Unlike these two algorithms, CFS and FCBF directly output optimal feature set out of the feature unit. This subset of features is expected to have the highest relevance and lowest redundancy. So in this case the entire subset is coupled with features of spectral energy under the same classifier and lead scenario for classification evaluation. Lastly, again the whole process iteratively applies to all lead scenarios and further all feature groups.

5. *Classification performance assessment*: Coupled with new features, spectral energy is subjected to classification for performance assessment. Following the coupling

strategy above, classification for ReliefF and InfoGain is done in order of high-to-low weighted features[3], and performance assessment is derived one by one. In regards to CFS and FCBF, one final performance is directly obtained as only one optimal subset is available.

Note that 10 runs of 10-fold CV is applied to obtain the classification performance assessment. Moreover, unlike what we did in Sec 5.3.1.2, no explicit trade-off analysis on SVM is made this time. This is because previously only spectral energy as one type of feature was considered, and no extra choices of features via feature selection were involved. That made it easy to perform trade-off analysis and demonstrate how trade-off was achieved. However this time situation is different. Here, a large number of potentially useful feature spaces are built and ready for assessment. Trade-off analysis in this case may not be practically necessary, as ultimately only the final assessment after optimisation for trade-off is of our interest. Thus, we perform optimisation of regularisation parameter C_s for SVM Case 2 by grid search (C = 2^{-5} to 2^{10}) without explicit discussion in this study. Optimised computational complexity with optimal number of support vector can then be achieved.

5.4.3.2 Experimental Results

Having explained the experimental procedure, it is possible to obtain the assessment of classification performance. Very importantly, only the subset of features from a feature unit that contribute to the highest accuracy is retained in the case of ReliefF and InfoGain; as for CFS and FCBF, the highest accuracy is directly available and thereby retained. This policy of retrieving highest performance estimation applies to situations under each classifier, each lead scenario and each feature group. As a result, we manage to obtain the highest accuracy each feature group can possibly augment spectral energy with via means of feature selection (A detailed description of the results and associated illustrative figures can be found in Appendix D). Moreover, since our final goal is to examine how many features from which feature group would contribute the most to augmenting spectral energy-based classification under certain lead scenario, we treat the classification results of our feature selection algorithms under the same lead scenario in an equal fashion. Because of this, we further extract these results (presented in Appendix D), and thereby leads us to Table 5.7 where only the highest accuracy among feature groups under each lead scenario and each classifier is retained. Apart from the accuracy, specificity and sensitivity are also given. More importantly, this table also presents which feature group that contributes to the augment and from this group the number of features that are involved. Thus, adding with the original number of spectral energy features, the total number of features required for classification can be derived. Also, for SVM the number of support vectors (SVs) is given, as it is required

[3]These weighted features can be found in units on the right hand side of Figure 5.4.

Table 5.7: Spectral energy-based classification performance augmented with more features, and the associated computational complexity for all five classifiers.

	Ld. Scn.	**Spe (%)**	**Sen (%)**	**Acc (%)**	**Fea. Grp.**	**# of Added Fea.**	**Total Fea.**	**# of SVs**	**NG ($\log_{10}$)**
LDA	**1**	95.96	75.96	85.96	Time	6	9	/	4.2265
	2	92.88	80.96	86.92	Time	7	11	/	4.3137
	3	93.08	85.77	89.42	Duration	1	9	/	4.2265
	4	90.96	88.65	89.81	DWT	1	10	/	4.2723
	5	95.58	87.20	**91.25**	Peak	6	14	/	**4.4184**
QDA	**1**	96.15	84.62	90.38	Time	2	4	/	4.6488
	2	92.69	88.65	90.67	Time	2	8	/	5.1732
	3	89.42	90.38	89.90	Time	3	10	/	5.3496
	4	86.92	93.65	90.29	Time	4	13	/	5.5609
	5	87.50	92.12	89.81	Time	3	12	/	5.4960
SVM$_L$	**1**	90.58	80.96	85.77	Time	3	5	56	5.8592
	2	90.96	85.77	88.37	Time	3	7	81	6.1301
	3	90.00	89.23	89.62	DWT	11	17	39	6.1398
	4	92.12	87.69	89.90	Peak	6	14	75	6.3487
	5	90.77	91.35	**91.06**	Peak	1	11	71	**6.2341**
SVM$_Q$	**1**	89.04	88.65	88.85	Duration	1	3	52	5.7590
	2	93.85	87.69	90.77	Duration	1	5	63	5.9691
	3	93.08	84.04	88.56	Time	3	10	65	6.1957
	4	92.88	87.88	90.38	Time	3	10	54	6.1152
	5	94.62	87.69	**91.15**	Peak	3	11	54	**6.1476**
k-NN	**1**	92.50	82.50	87.50	Time	2	3	/	5.7646
	2	98.08	83.08	90.58	Time	3	5	/	5.9766
	3	92.50	85.38	88.94	DFT	2	9	/	6.2252
	4	95.77	77.12	86.44	Peak	9	17	/	6.4975
	5	95.96	85.19	90.58	Time	1	11	/	6.3109

for computational complexity estimation. Finally, the computational complexity can be calculated as we did in Section 5.3.1.2.

To analyse Table 5.7, firstly we start with the classification performance. It can be seen that on average all classifiers show fairly high specificity, mostly higher than 90%. Meanwhile, they show comparatively low sensitivity, mostly ranging from 75% to 90%. In general, specificity is shown to be higher than sensitivity. This effectively means classifiers tend to label normal heartbeats more correctly than abnormal ones, due to the fact that the added features correlate more with the normal class than the abnormal class. As an overall measure of classification efficacy, accuracy lies around 90%, and in general it increases (LDA, SVM$_L$) or slightly fluctuates (QDA, SVM$_Q$, k-NN) along with lead scenario.

Secondly, it is not difficult to observe most of feature groups that contribute to the augment come from Duration, Peak, and mostly Time. DWT and DFT, on the other hand, bring in improvement to a better extent than the other three only under few lead scenarios and classifiers. Moreover, such observation actually proves that clinical features helps improving the classification performance (Section 5.4.1), as they closely relate to the clinical importance of the signals.

Thirdly, it can be seen that the number of features used for augment depend heavily on the lead scenario and the classifier. Importantly, these added features are selected as final outcome of four selection algorithms, and expected to derive the highest accuracy under their corresponding lead scenario and classifier. That is why improvement can be observed when comparing with Table 5.2 and Table 5.3. Improvement is seen to varying degrees. Especially, huge improvement ranging from 7% to 16% can be observed in 1 lead scenario for all classifiers. The whole pinpoints an important note here: in general, augmenting spectral energy features with more features facilitate our original classification in terms of performance. To say the least, introducing more features as the initial goal of this study is proven to be beneficial. In addition, the added number is generally quite small, making the associated total number to be generally less than 17.

Fourthly, as for the computational complexity, general growth can be easily seen in the transition from Table 5.2 and Table 5.3 to Table 5.7. Technically such growth is inevitable as more features are introduced. Particularly, though optimised, both SVMs still tend to exhibit higher complexity than before. Introduction of more features in hope of decreasing the number of SVs and further reducing complexity is shown to be infeasible. Instead, the number of SVs remains fairly high, and in some lead scenarios it is even higher than Case 2 in Table 5.3. In addition, on individual classifier perspective, gradual increase along with the lead scenario under each classifier can be seen. Notably, significant difference between classifiers can also be observed. That is, LDA enjoys substantially low computational complexity in all lead scenarios, and it also turns out to be the lowest. Specifically, one or even two order of magnitudes lower can be observed when compared to corresponding lead scenario under QDA and the rest, respectively. Overall, the entire observation signifies the low-complex capability of LDA under augment with more features for classification. This may further strengthen our observation and arguments in Section 5.3.1.2 as well.

Overall, it is not difficult to find that the highest accuracy that can be possibly achieved is about 91%. Three classifiers: LDA in 5 lead scenario, SVM_L in 5 lead scenario and SVM_Q in 5 lead scenario are seen to reach that level, and they are all demonstrated to exhibit slightly higher accuracy than the same scenarios but without augment (Table 5.2 and 5.3). This lead us to believe that, augment with more relevant features indeed contribute to the classification performance. Now, these three classifiers tie in terms of classification performance, leaving computational complexity to conclusively judge the most optimal classifier. From Table 5.7, it is fairly obvious that LDA significantly

Table 5.8: Simulation results of multiple heartbeat classification augmented with more features for all classifiers.

	LDA	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
Spe (%)	85.71	25.49	66.00	74.00	72.55
Sen (%)	89.13	54.35	89.13	86.96	89.13
Acc (%)	87.37	39.18	77.08	80.21	80.41

outperforms the other two by about two orders of magnitudes. That effectively signifies LDA enjoys much lower power consumption when labelling new sample. Thus far, LDA is considered to be the most optimal classifier.

5.4.4 Multiple Heartbeat Classification

Similar to Section 5.3.2, the procedure of running multiple heartbeat classification augmented with more features is discussed here. In this study, the total number of heartbeats for each lead and for each patient is also 7. More importantly, the heartbeat signals are exactly the same as previous experiment in Section 5.3.2, assuring direct comparison between the two. Besides, TDMG is operated to obtain the boundaries of P, QRS and T waves within each of the heart-betas. From there, spectral energy of interest can be derived. However, what is different to Section 5.3.2 is the inclusion of more features and the trained classifiers. On one hand, with the outcome of TDMG, features like peaks and durations of waves are derived and accordingly coupled with spectral energy to form new feature vector for each beat. On the other hand, due to the change of feature vector, trained classifiers from Section 5.3.2 are no longer of use. Instead, we deploy the trained classifiers from single heartbeat classification in Section 5.4.3.2, all of which are under their respective 5 lead scenarios. Following, multiple heartbeat classification can be operated and labels of the 7 heartbeats will be generated accordingly. The whole process can also be illustrated as in Figure 5.2. Last but not least, decision-making scheme for multiple heartbeat classification remains the same as Section 5.3.2, so do the evaluation criteria. Results can be found in Table 5.8.

In terms of specificity, it can be seen that LDA performs the best at 85.71% amongst the rest. Other achieves less to varying extent, having 25.49% for QDA, 66% for SVM$_L$, 74% for SVM$_Q$ and 72.55% for k-NN. As for sensitivity, all classifiers perform on fairly similar level except QDA. These two metrics signifies that, LDA is the best at correctly predicting normal and abnormal heartbeats for individual patients among all. This also sums up in accuracy, where LDA achieves 87.37% and is regarded as the best to the rest. In addition, when comparing to single heartbeat case (Table 5.7), performance reduction is observed. Similar observation was also found in Section 5.3.2. In this case, despite having augment of features in place, performance reduction is seen to be unavoidable. So it is likely that heartbeat we picked for classifier training may not be sufficiently

representative. Nonetheless, the good side of it is that, improvement of 2% to 11% for all classifiers in multiple heartbeat classification can be observed when compared with Table 5.4, only except QDA. From there it reveals the beneficial influence of augment of more features on multiple heartbeat classification.

5.4.5 Discussion

Further extending our study of spectral energy based classifications in Section 5.3, spectral energy augmented with more features has demonstrated to surpass those without augment in single and multiple (except QDA) heartbeat classification in terms of classification performance in general. This in fact testifies our point of introducing more features in Section 5.3.3. Yet, such improvement was commonly done at the expense of computational complexity. Despite of this fact, through the investigation LDA has continued to render itself the most optimal classifier, by leveraging its advantage in complexity while keeping the classification accuracy as high as complex classifiers can do. Specifically, 6 features from feature group Peak contribute to such enhancement (Table 5.7).

5.5 Concluding Remarks

As a further investigation of classification efficacy for spectral energy, this chapter has mainly covered two scenarios: spectral energy-based classification and spectral energy-based classification augmented with more features. Both of them have been experimentally carried out by considering single heartbeat classification and multiple heartbeat classification together. Findings can be concluded in the following. Firstly, as for classification performance, in general slight improvement from spectral energy-based to augmented spectral energy-based classification were observed in all lead scenarios in single heartbeat classification, particularly huge improvement ranging from 7% to 16% was seen in 1 lead scenario for all classifiers. Same happened to multiple heartbeat classification, where 2% to 11% improvement was observed except QDA. Improvement on both single and multiple heartbeat classifications could be possible thanks to the effective use of feature selection algorithms. Secondly, as for computational complexity, specific optimisation for SVM in regards to trade-off between classification performance and computational complexity was made through the study. In spite of this, SVM together with QDA and k-NN were seen to exhibit one or two order of magnitudes higher complexity than LDA but with no superiority in classification performance. In other words, LDA have shown excellent balance between the two important metrics, making it the most optimal classifier among them all. However, if we specifically look at LDA in 5 lead scenario, both its single and multiple heartbeat classification performance in spectral energy-based classification augmented with peaks shows only ~2% improvement

compared to those without, which is ultimately not of enormous amount. Despite of a small extent, we have justified the goodness of deployment of more features in other aspects, like 1 lead scenario and multiple heartbeat classification. Now, as for hardware implementation, we aim for high-performance and yet simple design. More importantly, it serves as a demonstration of the concept of spectral energy-based classification. As a result, finally we choose to implement LDA in 5 lead scenario without peaks as our hardware solution. More will be covered in the next chapter.

Chapter 6

Hardware Architecture for On-Body ECG Classifier System

In last chapter, we have justified the benefits of introducing more features to enhance the spectral energy-based classification. LDA in this case revealed its advantages in both classification performance and computational complexity. After thorough comparisons, LDA in 5 lead scenario was chosen. However, despite augmented with features from feature group Peak, little improvement of classification performance was observed. On this basis, it prompted us to choose LDA in 5 lead scenario without augment for hardware solution in that (1) 2% of improvement may not be well-worth as extra hardware would be required to handle those features in terms of feature generation and sample prediction; (2) the concept of spectral energy-based classification can still be demonstrated even without enhancement of features. Therefore, in this chapter we presents the associated hardware architecture of LDA-based on-body classifier for normal and abnormal ECG classification.

This chapter is organised as follows[1]: Section 6.1 gives a broad view of the system regarding how the system works as a whole, main functionality of each sub-block as well as the signal data flow throughout the process. Section 6.2 covers the detailed architecture and implementation of the dedicatedly functional sub-blocks of our design. Following, Section 6.3 discusses the implementation set-up, verification and performance analysis regarding total dynamic power, cell area, throughput, etc. In the end, Section 6.4 concludes the chapter.

[1]The contents of this chapter have partly appeared as "Design of a Low-Power On-Body ECG Classifier for Remote Cardiovascular Monitoring Systems" by Chen *et. al.* in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems.*

6.1 Overview of the System

In light of findings from Chapter 5, we have decided to implement a low-energy VLSI solution for LDA in 5 lead scenario, specifically with the view of integrating it on a body-worn ECG sensor enabling low-energy ECG classification. Such investigation will allow us to comment on the implementability of the proposed classifier on resource constrained ambulatory ECG sensors, used in remote CVD monitoring systems. The reason for implementing the system as an ASIC instead of a microprocessor-based or DSP solution has twofolds: it is well established that a general-purpose processor or a DSP-based design for any application consumes at least 2 to 3 orders more power compared to its equivalent ASIC implementation; and an ASIC design is far more easier to be integrated with the ECG sensor, particularly in body-worn sensor networks. As power consumption is the main constraint in our case, it is preferable to choose an ASIC solution. Thus, in terms of resource usage, we aim to design the system using minimal number of multiplications. Haar basis function used for DWT analysis in our case already results into low arithmetic complexity. In particular, we eliminate the square root and division operations involved in Haar based DWT cD coefficient computation by combining the processes of coefficient computation and the squaring operations for the spectral energy estimation (Section 6.2.3). For the sake of presentation, here the following conventions of notation are used: ${}^{1}\mathrm{T}_{5}$ represents T wave boundary of lead 1 at level 5, while ${}^{2}\mathrm{QT}_{345}$ indicates QT wave boundary of lead 2 at level 3, 4, 5 and so on.

The block diagram of the architecture is illustrated in Figure 6.1. According to our previous discussion, the final implementation of the system explicitly serves the 5 lead scenario of the LDA classifier. Firstly, the raw signal (Sig) is fed into the DWTLVm blocks. DWT cD coefficients of decomposition levels 2-5 are then computed using the DWTLVm blocks in parallel, where m indicates the index of decomposition level. The corresponding DWTLVm blocks are selectively activated when necessary. The activation depends on which lead is under consideration, since the features associated with the leads may belong to any of these decomposition levels according to Table 4.6. Secondly, using the appropriate cD coefficients from this block, the corresponding spectral energies can be computed in the CoefSelection&Squaring block (CSS). The entire process is done in a lead-by-lead sequence so as to repeatedly take advantage of these functional blocks in our system. We intentionally use a sequential approach here since, according to the clinical specification, classification of normal and abnormal ECGs within a few seconds time is very much acceptable. Therefore, although the architecture could be made faster by parallelising the functional blocks and dedicating them for each of the leads, it will eventually result in over-engineering and also will have a detrimental effect on the overall energy consumption. Thirdly, once all desired spectral energy features are obtained, feature vector can then be produced in FeatureVectorGenerator block (FVG). It is then followed by LDA block which simply run the LDA algorithm to achieve the

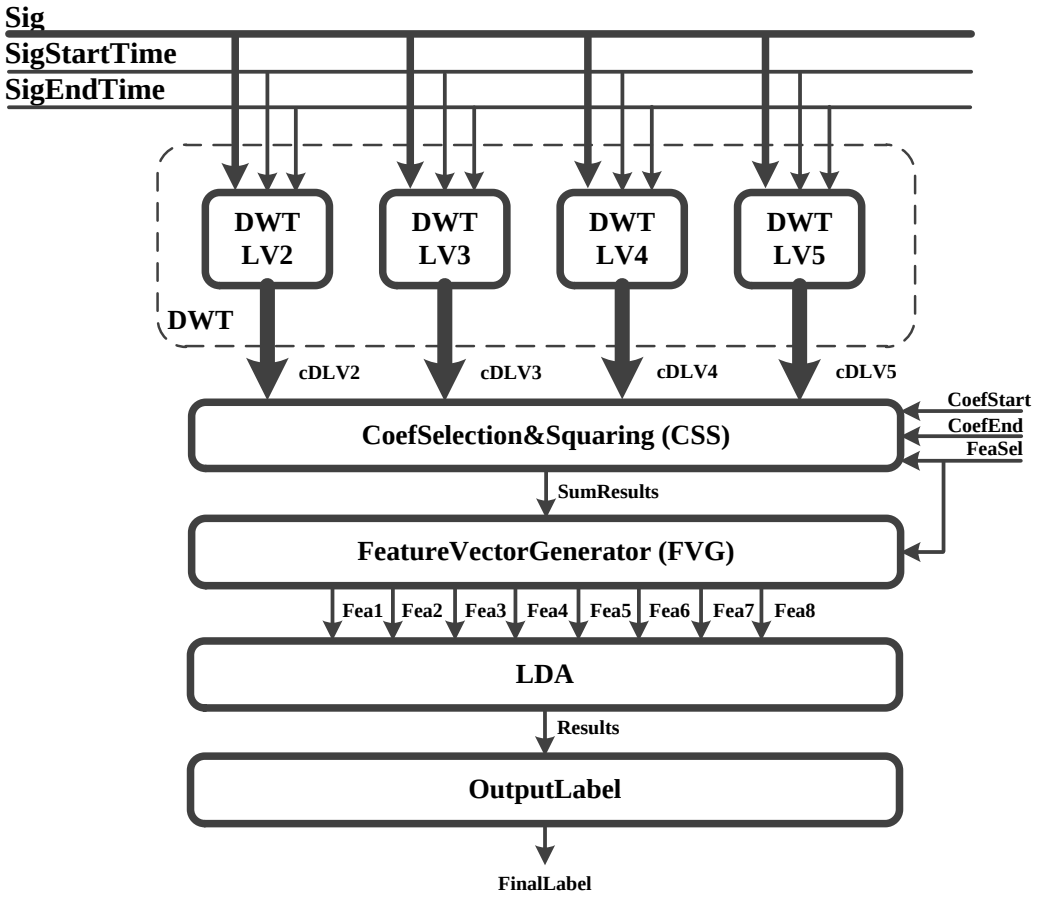


Figure 6.1: Hardware architecture of the system.

final labelling results. Finally, this is then passed to the OutputLabel block in order to output the final label of this single beat.

Two signals – SigStartTime and SigEndTime are used to indicate the start and end time of one heartbeat, and hence the time window of the signal within which the DWT coefficients will be computed. Importantly, as our focus is mainly on classifier implementation, no specific implementation of ECG fiducial point detection is considered. Therefore, two assumptions are made: (1) we assume that the fiducial point detection for the ECG waves is done by some other blocks outside of the present classification system; (2) we also assume that the corresponding wave boundaries of ECG beat components ($^{1}T_5$, $^{2}T_5$, $^{2}QT_{345}$, $^{3}T_5$, $^{3}QRS_2$, $^{3}QT_{35}$, $^{7}QT_{345}$, $^{8}P_5$) are directly available to the system once we have ECG signal at the input to the present system. CoefStart and CoefEnd provide the CSS block with the signals that indicate the wave boundaries detected externally. Also, they are used for selecting the set of appropriate cD coefficients for spectral energy computation, with FeaSel indicating which specific spectral energy CSS should compute (refer to Table 4.6), as well as the feature vector FVG to generate. CoefStart, CoefEnd, FeaSel are signal generated from the controller of our

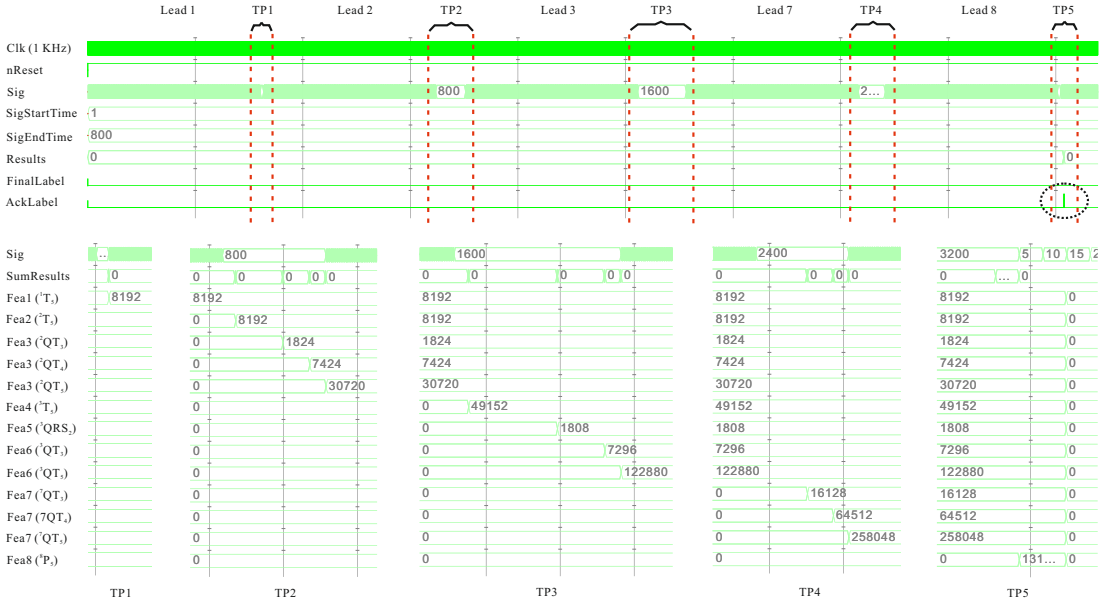


Figure 6.2: Main data flow of the design.

systems, which is not shown here. Note that FeaSel is a select signal that selects one of 13 cases of the spectral energy of ECG beat components considered in this work.

The associated data flow for information processing is depicted in Figure 6.2. To illustrate the flow with an example, assume that the input ECG signal (Sig) feeds in random samples with a length of 800 for each lead. SigStartTime and SigEndTime are set to 1 and 800 respectively as the time frame of the signal. Since the ECG PQRST complexes for each lead is fed sequentially, the DWT coefficients at selective decomposition levels are computed in parallel and on-the-fly. The coefficient squaring operation required for computing spectral energy at each level is done in the period TP (TP1 to TP5), followed by the sequential addition operation to compute our spectral energy features as a whole (particular for those which require spectral energy from more than one DWT decomposition levels). Once the entire spectral energy computation is finished, AckLabel is asserted in time period TP5, indicating the FinalLabel is available (which is '-1'[2] in this case). The process is shown in the upper half of Figure 6.2. The lower half of Figure 6.2 explicitly shows the timing details of when exactly each summation of the squared coefficients is done for each lead in TP1 to TP5. These values are stored temporarily into an intermediate register bank and keep on accumulating along the time from TP1 up until TP5. These are then used by the FVG block to generate the feature vector. From there, LDA block receives the feature vector. Eventually, the final classification label is produced in LDA for the current single heartbeat classification during TP5.

To process multiple heartbeat classification in real-time, the system requires an additional memory bank to store the multiple heartbeats. This is because our system is based

[2]FinalLabel is in the form of 1 bit 2's complement.

on single heartbeat classification and thus needs to be applied iteratively for labelling individual single heartbeat classification scenarios.

6.2 Design of Sub-Blocks

In this section, details of each block comprising the overall system are described. Illustrative diagram of the blocks are shown in Figure 6.3.

6.2.1 DWTLVm Block

In accordance with DWT filter transfer function (Equation 3.9), 2^m consecutive data samples of the ECG signal are used to compute the detail coefficients cD at each decomposition level, where m = 2, 3, 4, 5 in our case. Although the transfer function involves the term $\frac{1}{\sqrt{2}}$ when calculating the coefficients (Equation 2.11), it can be ignored in this block and then be taken into account as shifting later in the FVG block where features are generated. Doing so in turn simplifies the DWT computation block. Therefore, the DWT coefficients can be computed using simple additions and subtractions. This is done on-the-fly with the incoming signal samples. The architecture for generic coefficient computation at the decomposition level m is shown in Figure 6.3(a).

6.2.2 Coefficient Selection and Squaring Block

Coefficient Selection and Squaring (CSS) Block is the one that receives the detail coefficients and produce the spectral energy of a specific ECG wave at a certain DWT decomposition level. Its block diagram is shown in Figure 6.3(b). Four register banks, namely cDLV2RegBank, cDLV3RegBank, cDLV4RegBank, cDLV5RegBank are used for storing the detail coefficients generated by the DWTLVm block. Depending on which lead we process, one or more of the register banks are utilized to store the coefficients of certain levels that are necessary in squaring operation. The others will be temporarily ignored during the time of processing that particular lead signal. Once the expected coefficients are successfully stored in the associated register banks, FeaSel is asserted and the corresponding cDStoreSel signal is generated via Select Signal Converter for selecting appropriate coefficient register bank(s) for spectral energy computation. A synchronous up-counter is used in Squaring block, with CoefStart and CoefEnd indicating the start and end values of the count operation upon the selected register bank. Henceforth, the coefficients from the CoefStart up to the CoefEnd will be sent to the Squaring sequentially for squaring operation, followed by Accumulator which sequentially sums up squared results. Eventually the overall sum of the selected squared coefficients is outputted as SumResults to the next block. Note that, since the timing requirement of our

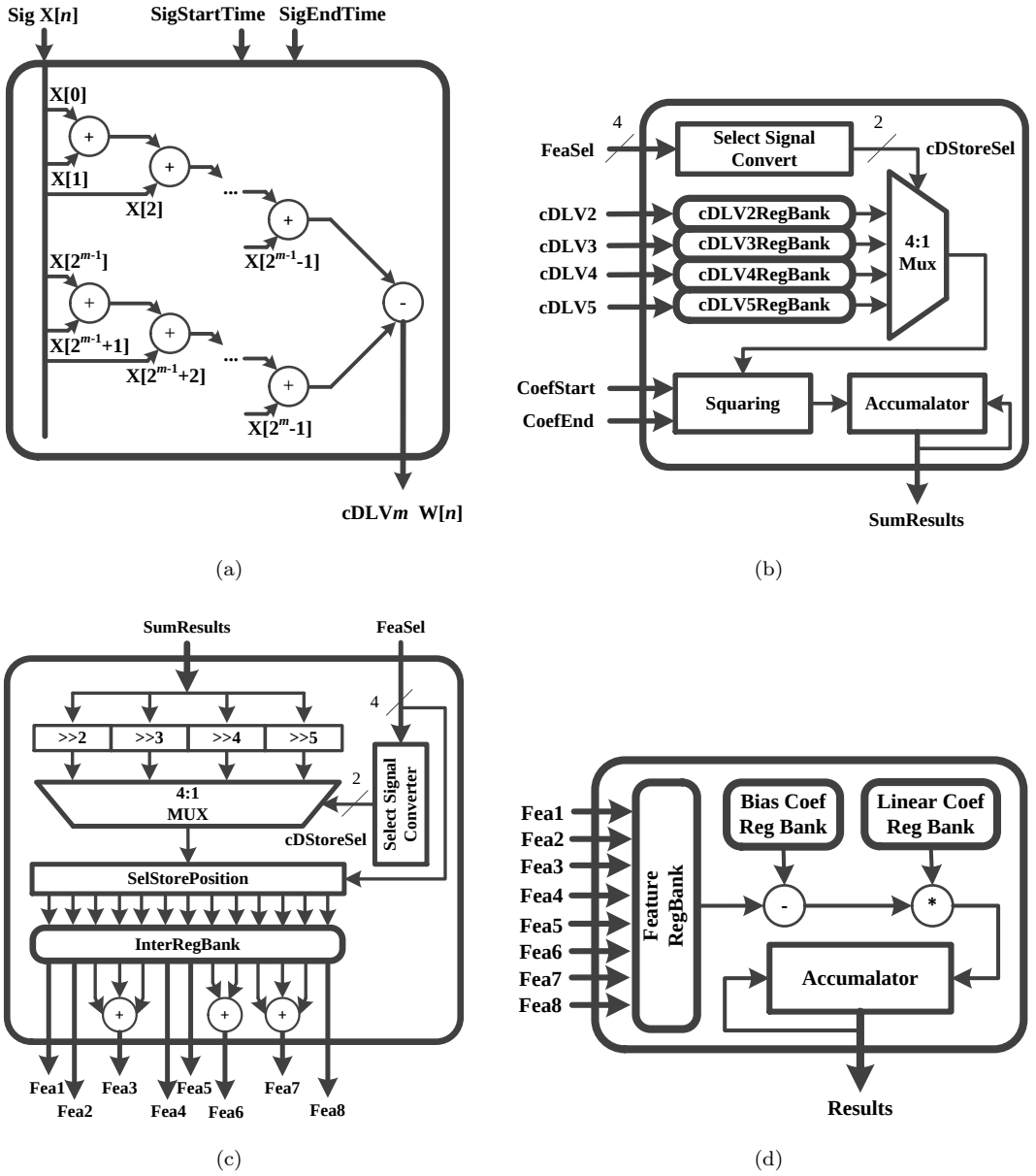


Figure 6.3: Hardware architecture of the sub-blocks: (a) DWTLVm; (b) CSS; (c) FVG; (d) LDA.

system can tolerate longer processing time, CSS block is implemented in such a way that it only requires one multiplier and one adder. With such, the squaring and summation of the squared coefficients, and thus the spectral energy, can be computed.

6.2.3 Feature Vector Generator Block

As we previously discussed in subsection 6.2.1, $\frac{1}{\sqrt{2}}$ term in the Haar DWT transfer function was temporally removed. However, to maintain the computation precision while reduce the computational complexity of our system, $\frac{1}{\sqrt{2}}$ term has been implicitly

considered not in the stage of detail coefficient computation, but in the stage of spectral energy computation. To better explain it in details, Equation 6.1 shows the spectral energy calculation procedure of the first squared coefficient at level 2. It can be clearly seen that, at the end of the equation, there is a squared coefficient along with the squared term $W_2[0]^2$. This coefficient is resulted from the previous $\frac{1}{\sqrt{2}}$ term in cD coefficient computation. After squaring, ultimately the $\frac{1}{\sqrt{2}}$ term will lead to $\frac{1}{4}$, which can be considered as simple shifting. Same principle can be applied to the decomposition levels 3, 4, 5 and thereby reducing arithmetic complexity of the system further.

$$\begin{aligned}
&\left[\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}X[0]+\frac{1}{\sqrt{2}}X[1]\right)-\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}X[2]+\frac{1}{\sqrt{2}}X[3]\right)\right]^2 \\
=&(\frac{1}{2})^2\left[(X[0]+X[1])-(X[2]+X[3])\right]^2 \\
=&(\frac{1}{4})W_2[0]^2
\end{aligned} \tag{6.1}$$

Figure 6.3(c) shows the architecture of the FVG block. The above mentioned shifting process is implemented as the first part, with an aim of adjusting the SumResults from the CSS block. Following the shifting, there is a multiplexer associated with cDStoreSel as the select signal which again is produced from Select Signal Convert with FeaSel as input. Here, this multiplexer is to select the proper shifted SumResults out of the four shifted ones. In order to store it into the InterRegBank at the proper position so that the final features generated are in correct order, SelStorePosition assists to localise the position in InterRegBank according to the FeaSel. Finally, once all the adjusted squared SumResults are stored appropriately, operation of outputting the features will be initiated. Before the series of Fea are sent out, Fea3, 6 and 7 require one more summation to derive the final result from the corresponding values in the register bank while the rest are sent straightaway to the LDA block.

6.2.4 LDA Block

Mathematically, LDA consists of a set of coefficients of linear function and one constant (Equation 2.22). These coefficients and constant have already been derived in our simulation (single heartbeat classification analysis in Section 5.3.1) using Matlab. However, since the constant term turned out to be too small (approximately 10^{-15}), we ignore it in our LDA block. Therefore, only LinearCoefRegBank is used to store the set of coefficients. In addition, standardisation (Equation 2.12) should also be done on each feature before operating LDA. To simplify this process in our design, the mathematical issues have been taken into account before the implementation of our design. That is, as we are only concerned about the sign of the results of LDA, the division over standard deviation can then be eliminated by multiplying a dedicated constant on

both side of the linear equation of LDA and thus cancelling out the denominator (i.e. standard deviation) without changing the polarity of the final results. To maintain the accuracy, linear coefficients of the equation are altered correspondingly. Substraction with associated mean value from each feature is also adjusted accordingly. Henceforth, BiasCoefRegBank is only needed to store these adjusted mean values. With all these values ready, Fea1-8 generated from the previous block are firstly subtracted by the corresponding adjusted mean value in sequential order, followed by multiplication with the associated coefficient of the linear function. The result is accumulated with the previous one. When it is done, the final output Results is produced. Figure 6.3(d) shows the hardware implementation of the LDA block.

6.2.5 OutputLabel Block

This block is used for simply extracting the sign bit of the final output generated from LDA block as the label. This label is categorised into two classes : >0 represents abnormal and <0 normal.

6.3 System Implementation and Verification

The proposed architecture (see Figure 6.1) was coded in Verilog and Synopsys Design Compiler was used to synthesize the HDL code at 1 KHz clock frequency and 1.08 V supply voltage, using the STMicroelectronics 130 nm technology library. The power consumption[3] and total cell area as well as area after Place and Route (PnR) of the entire system can be seen in Table 6.1. For ease of comparison between different technologies, NAND2 equivalent area (gate count) is also given. Similarly, Table 6.2 shows the power consumption, cell area and NAND2 equivalent area for specific modules within the system. In particular, as the main classification functional module, LDA block consumes an estimated 17.04 nW. In total, the power consumption of the design was estimated at 182.94 nW. From both tables, we conclude that the proposed classification architecture is ideal for implementation in low-power mobile CVD platforms and also has the potential to be integrated with an ambulatory ECG sensor, in the form of a standalone ASIC. Figure 6.4 depicts the core layout with labeled blocks after PnR. The associated area is 0.979 mm^2 and the equivalent NAND2 area is 161.8K. The reason why these measurements are larger than the area in Table 6.1 is due to practical considerations as cell placing and signal routing are reflected in PnR.

In addition, to obtain the throughput of the design, we consider the time instance when the first ECG sample is fed into the system as the starting point up to the point when

[3]Power estimation was done in Synopsis PrimeTime.

Table 6.1: Synthesis results for the proposed system.

Technology	ST130nm
Global Operating Voltage	1.08 V
Clock Frequency	1 KHz
Total Dynamic Power	182.94 nW
Total Cell Area	0.701 mm^2
Total Cell Area (PnR)	0.979 mm^2
NAND2 Equivalent Area	115.5K
NAND2 Equivalent Area (PnR)	161.8K

Table 6.2: Synthesis results for specific modules.

Module	Dynamic Power (nW)	Area (mm^2)	NAND2 Equi. Area
DWTLV2	87.49	0.291	48.9K
DWTLV3	44.36	0.152	25.1K
DWTLV4	23.02	0.080	13.2K
DWTLV5	12.92	0.046	7.6K
CSS	3.45	0.093	15.4K
FVG	6.90	0.026	4.3K
LDA	17.04	0.026	4.2K
Controller	0.34	0.003	0.5K

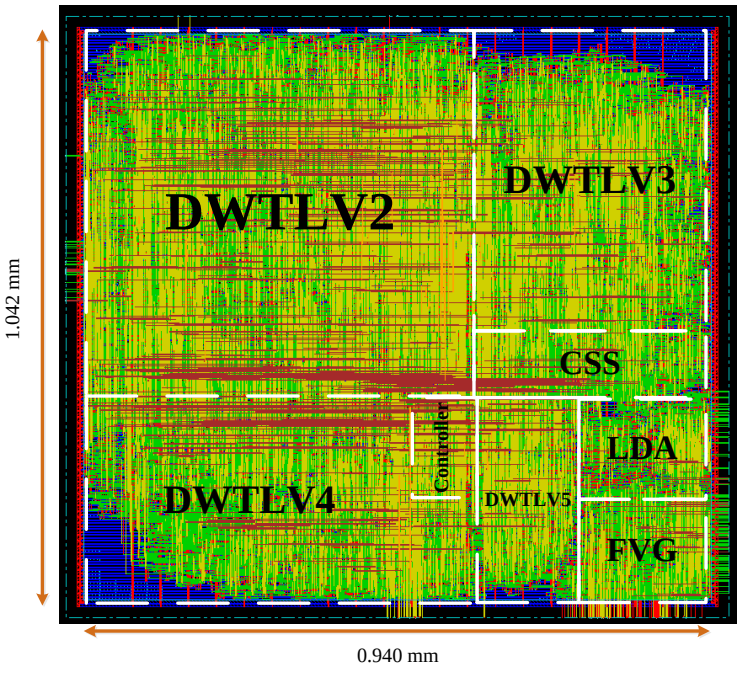
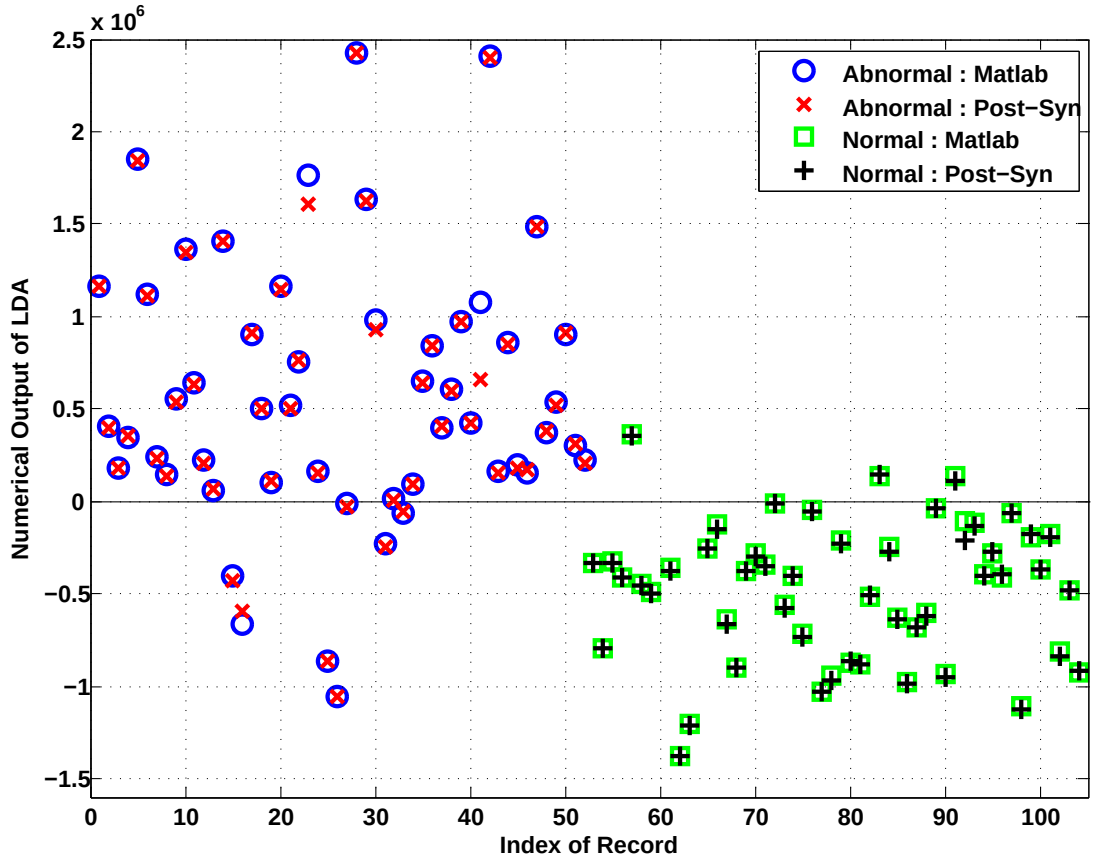


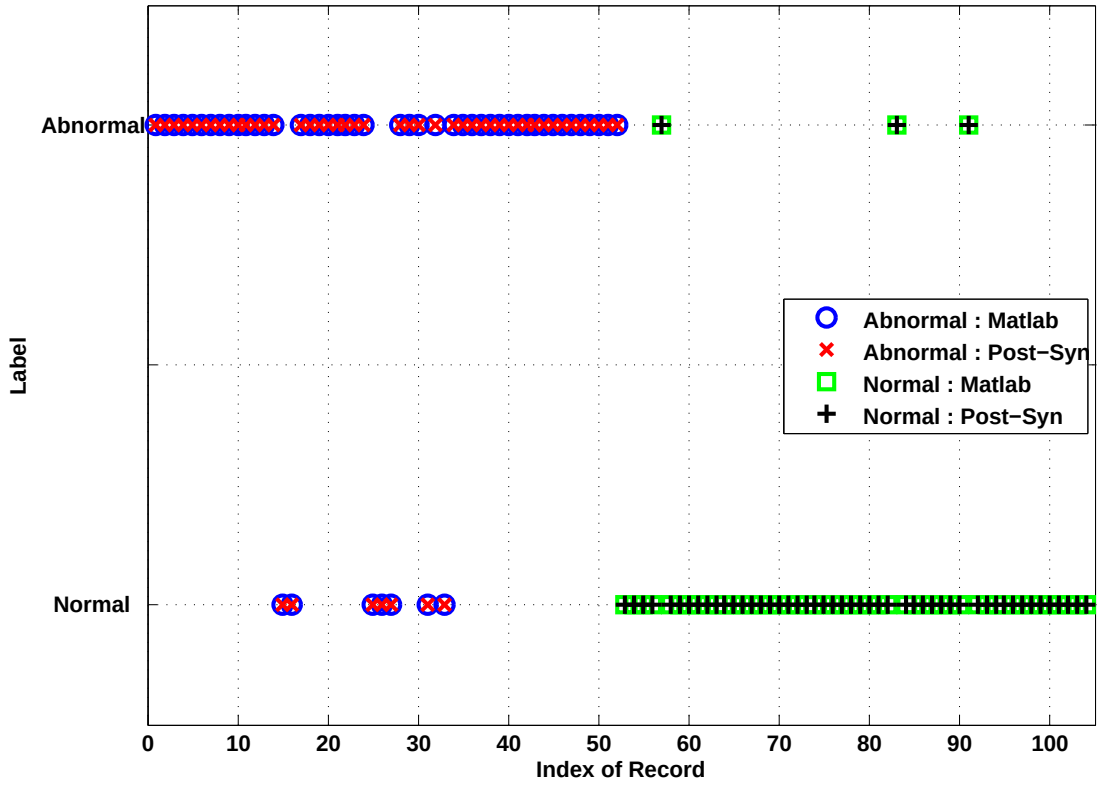
Figure 6.4: Core chip layout.

the final label is produced. The throughput will vary depending on the number of samples of the heartbeat and also the duration of the ECG beat components, in terms of samples, used in calculating the spectral energy features. To provide an approximation of the throughput, a single heartbeat classification was conducted where the length of the heartbeat has been set to 800 samples and the boundaries of P, QRS, QT, T were set to 128 samples, 96 samples, 416 samples, 160 samples respectively. When considering an ECG sampling frequency of 1 KHz (the sampling frequency of the PTBDB), these values are within the normal clinical limits for these parameters. Under this set-up, 4516 clock cycles (at 1 KHz operating frequency) are required to output a label, which is approximately 4.5 s. Such latency in classifying the ECG signal and to that extent, trigger the *danger* alarm in remote CVD systems is well accepted by cardiologists and physicians. This fact, in combination with the detrimental effect on the power consumption, prompted us not to consider a parallel architecture in our design.

In addition, a thorough experimental verification is conducted to compare the classification results obtained from the VLSI system at post-synthesis level against the Matlab-based implementation. We consider a single heartbeat classification where 5 leads are available with LDA as the classification method. The 104 records (52 abnormal and 52 normal), described in Section 4.3.1 are again utilized for this experimentation. Initially, we use all 104 records in order to train the LDA classifier. For simplicity, this operation take place in Matlab environment. Once the LDA coefficients are derived, they are imported in the synthesized core to define the LDA parameters. In the testing phase the same 104 ECG records are utilized. A Verilog-coded testbench is constructed to define the testing vectors (104 records) that are used as the input data to the synthesized system. Ultimately, we compare the numerical value of the trained LDA output produced by Matlab-based implementation to the synthesised system's. To illustrate the results, Fig. 6.5(a) shows the numerical value of the LDA output for both the Matlab-based and the synthesised system for each testing vector. From there, it is evident that the values from the two either match or are very close. Fig. 6.5(b) shows the final classification label for each record based on the operation of the LDA classifier. It can be seen that the classification labels between Matlab and the synthesised system fully agree, even for the misclassified ones. In total we observed the same 7 abnormal records misclassified as normal and the same 3 normal misclassified as abnormal out of the total 104, in both implementations. This results in the following values for specificity, sensitivity and overall accuracy, Spe = 94.23%, Sen = 86.54% and Acc = 90.38% of the proposed system. This investigation fully validates the synthesised design of the proposed classification system. In essence, the multiple heartbeat classification scenario is also validated here, since it is simply an iterative application of the single heartbeat method on multiple heartbeats. Finally, due to the fact that ECG signals considered in our study are actual medical records, we expect the same level of performance in the real-life application of the proposed system.



(a) Numerical value of the LDA output



(b) Class labels produced for 104 records

Figure 6.5: The comparison between the Matlab and Post-synthesis implementation of the proposed classification system.

6.4 Concluding Remarks

In this chapter, a fairly power-efficient on-body ECG classification system has been presented. It is a system mainly comprising the feature generation of spectral energy as well as the LDA-based classifier, in a hope to demonstrate the feasibility of spectral energy-based LDA classification for classifying normal and abnormal ECG. Thanks to the simple mechanism of the classification model and the specifically efficient design of our sub-blocks, the entire system has shown to exhibit efficient power consumption with 182.94 nW (17.04 nW for LDA block) at 1 KHz & 1.08 V. Meanwhile, the classification performance have fully matched with our Matlab-based implementation, reaching 90.38% accuracy in classification.

Chapter 7

Conclusions and Future Work

Being proactive instead of reactive has been seen to be the future of our healthcare monitoring. To accomplish such goal, mobile healthcare monitoring system has come into the picture and proposed to deliver better quality of care and professional services to all stakeholders within this ecosystem. There, continuous monitoring has been the key of achieving the notion of proactive, as continuous variability analysis of the physiological data could help predict the impending vital event for the patient. However, traditional way of using on-body sensors for long-term transmission in mobile monitoring system is deemed to fail in maintaining continuous monitoring. This is because those sensors are hugely constrained in energy supply and computational capability under resource-constrained environment. Even so, most of the energy is consumed by the embedded radio front-end module for continuous data transmission, making it not sustainable in continuous monitoring. To tackle this problem, we have proposed to reduce the usage of the transceiver by finding abnormality of the ECG heartbeats on sensor, and generating an alarm and transmitting data via transceiver only when necessary. This significantly negates the requirement for the continuous use of the transceiver, in turn saving considerable amount of energy. An investigation into a signal processing unit (Figure 1.1) that is capable of extracting useful ECG clinical features, classifying normal and abnormal ECG heartbeats with robust feature, and associated hardware implementation has been made, and statistical analysis coupled with robust feature and potential improvement of classification performance with more features have also been discussed. As a closing chapter, Section 7.1 summarises the main points of research contributions throughout the thesis. Section 7.2, on the other hand, points out the potential research directions and challenges that may extend our current research effort to a higher level.

7.1 Thesis Contributions

In Chapter 3, two automated feature detection algorithms for ECG fiducial points have been proposed and developed: one is Time-Domain based Morphology and Gradient algorithm (TDMG) and the other is Hybrid Feature Detection Algorithm (HFDA). As for TDMG, morphological characteristics of the ECG waves in time-domain have been used together with gradient-based analysis so as to extract the clinically important fiducial points. Specifically, zero-phase bandpass filter coupled with moving average smoothing technique has been applied to filter noise-corrupted ECG. After that, gradient sequence which is required during the analysis has been created by moving slope filter. Having obtained noise-clear ECG and its gradient sequence, initial estimation of the points were achieved. Following that refinement was done, with an aim of improving the initial results from last stage. Notably, predefined searching window and adaptive threshold policies have been commonly deployed in these two stages. On the other hand, as for HFDA, both frequency- and time-domain analysis via DWT have been utilised to extract the same information. Similar procedure of feature detection has been applied, but with different predefined searching windows and adaptive threshold policies. Finally when came to validation, it has been shown that both algorithms achieve relatively poor results with respect to the state-of-the-art algorithms, except a few features (Table 3.4 and Table 3.5). Despite so, spectral energy (as discussed in next Chapter) as a feature exhibits robustness against feature misdetection. This in turn effectively compensate the poor performance of the feature detection algorithms. Besides, between our two proposed algorithms, results have shown that TDMG performs relatively better than HFDA in detecting QRS and T boundaries except P boundaries. Since the corresponding fiducial points are important to our spectral energy-based ECG classification in Chapter 4, we have opted to use TDMG as our primary ECG feature detection tool for later experiments.

In Chapter 4, firstly four set of derivation of spectral energy following different strategies have been examined against our database. Specifically, DFT, DWT and thresholding policy have been used individually in each set upon either entire ECG heartbeat or specific wave components of the ECG heartbeat. To evaluate and compare the effectiveness of these four approaches under their lead scenarios, five selected classifiers namely LDA, QDA, SVM_L, SVM_Q and k-NN have been exploited to provide a platform of classification judgement. Eventually, Set 4 that utilised spectral energy of the ECG wave components via DWT has shown to outperform the other sets in terms of classification performance (see Table 4.1, 4.2, 4.4, 4.7), and thus been chosen as our primary method for extracting spectral energy from ECG heartbeat. In addition, to further justify the goodness of spectral energy, its robustness against misdetection error introduced by ECG feature detection algorithm has been investigated. By applying artificial error injection, statistical analysis of the variation of spectral energy as well as classification performance analysis using spectral energy as a feature, under worst-case misdetection error,

could then be possible (see Table 4.10, 4.11 and 4.12). Ultimately, we have found that high-frequency and combined frequency group of spectral energy performs remarkably well against misdetection with all classifiers in both analyses, whereas low-frequency group of the same kind exhibits unacceptable deviation from the ground truth in the first analysis and partly acceptable results in the second.

Given the conclusion from Chapter 4, it is seen that there exists scope for improvement of classification. Among different methods, it is possible to augment the spectral energy-based classification by adding some useful features. In Chapter 5, prior to investigating the effect of more features, spectral energy-based classification has been re-visited. It has been done as to single heartbeat classification where specific care of trade-off between classification accuracy and computational complexity has been taken specifically for SVM (see Table 5.2 and 5.3), and multiple heartbeat classification (see Table 5.4). Following, spectral energy-based classification augmented with more features has been visited. Potential features generated via Time domain, DFT domain and DWT domain have been collected and fed into our experimental scheme of selecting most relevant features (Section 5.4.3.1). This scheme involves four widely used feature selection algorithms — ReliefF, InfoGain, CFS and FCBF. From there, the best selections of feature have been selected according to our classifiers and their lead scenarios. Similarly to the prior investigation, single heartbeat classification and multiple heartbeat classification (see Table 5.7 and 5.8) have also been carried out this time, with an aim of studying the improvement in both scenarios. Ultimately, improvement of classification performance has been observed in general in both single and multiple heartbeat classification. On the other hand, it has been seen that LDA has shown one to two order of magnitude lower computational complexity than the rest of the classifiers, while having competitive classification performance in high lead scenario, especially 5 lead scenario. Both observations have made LDA the best candidate in balancing classification accuracy and complexity.

Since only ~2% improvement was observed in LDA when augmented with more features, we have opted to design and implement ASIC solution for LDA-based 5 lead scenario under spectral energy-based classification without such augmentation. Therefore, in Chapter 6 we have explained the hardware architecture of the design and its functions from a broad view point, followed by detailed discussions on each sub modules. Finally, implementation and verification of the system have been presented in Table 6.1, 6.2 and Figure 6.5. Overall, the entire system consumes 182.94 nW at 1 KHz clock frequency and 1.08 V supply voltage using ST 130nm technology, while achieving 90.38% classification accuracy.

Above all, this thesis has mainly contributed to the enhancement of diagnostic quality of ECG in mobile monitoring environment, both from signal processing, machine learning and digital system perspective. The investigations carried out throughout the

thesis were substantiated by a number of experiments and comparisons involving different database. It is hoped that the findings of our research will contribute towards more effective and efficient signal processing, and intelligent and robust classification for ECG applications as well as low-power solution for ECG classification in mobile environment, and eventually better quality of service delivery for ECG-based mobile healthcare system.

7.2 Future Research Direction

The following have been identified as our potential future research challenges during the research study undertaken in this thesis.

7.2.1 Decision-Making Schemes on Multiple Heartbeat Classification

Currently the decision-making scheme for classifying multiple heartbeats in Chapter 4 is very simple: simply by voting, the higher scores of normal/abnormal heats will be output as the final decision. But when ectopic beats occur, in practice the number of abnormal beats may not necessarily be larger than the normal ones. So, according to the scheme above, one may classify the patient as normal. However, in practice, having abnormal beats less than normal beats in an *on and off* manner may indicate abnormal episode of the patient. So, possible solution to this issue is to develop an elaborate decision-making scheme that can reasonably utilise the predicted labels of the heartbeats. For example, if one abnormal beat was found, the scheme should be able to see how many abnormal beats are in the vicinity. Depending on the threshold we set, the patient might be classified as abnormal once the number of vicinal abnormal beats goes beyond the threshold.

7.2.2 Investigation into Integrating Both Feature Construction and Feature Selection for Efficient Feature Generation

In machine learning, feature extraction comprises of feature construction and feature selection. Usually, these two processes are operated separately. Even though feature construction may generate non-relevant features, researchers tend to use feature selection algorithms to filter out those non-relevant ones afterwards anyway, as operations are normally carried out on high-capability computers and optimal results will be achieved very easily.

In general, feature construction in signal processing (e.g. feature detection of ECG fiducial points in Chapter 3) is designed to deliver useful features, but with no guarantee

of being relevant to the sample classes. Particularly, if the feature construction algorithm was expected to produce huge number of features, chances are part of them may not be totally relevant to the classes of the samples, leading to impaired classification performance in the end. Spending huge effort in delivering non-relevant features to classification is not well suggested. So, to avoid such predicament, potential solution may lie in integrating feature selection process into feature construction. Possible methodology might include putting well-chosen feature section algorithms in the middle of the design process of feature construction algorithm. Before finalising the output of the algorithm, these outputs may be evaluated by feature selection techniques. If the output was not relevant to the sample classes, it may have not be implemented and decision has to be made by the designer about preserving this output or not. Overall, it is hoped that such integration can facilitate the process of effective feature construction. In particular, computational complexity of producing features is of big concern in mobile applications. Since achieving low complexity is the main goal, it may be beneficial if we only produce those relevant features in the first place, therefore preserving energy and computational resources for longer operation time.

7.2.3 Exploration of Potentially Useful More Sophisticated Classifiers

Focusing only on the single heartbeat classification in Section 5.3.1, classifiers like SVM, k-NN in fact performs better than LDA in 1-2 lead scenario in terms of accuracy. This in turn leads us to think deeply: 1) what if in some cases where 1-2 number of leads are strictly required (e.g. ICD), or 2) what if maximal accuracy is highly demanded and considered as the most important factor in mobile CVD monitoring system? Such situations would call for more advanced and sophisticated classification model to accomplish the tasks. However, practical problems for these classifiers will then emerge as being highly computationally complex, hence high power consumption.

To overcome this problem, potential solutions might lie in optimisation of classification algorithm as well as the associated hardware architecture. In this case, optimisation could reduce overall arithmetic computations by reducing division and multiplication at the expense of shifting and addition. In regards to hardware architecture, Distributed Arithmetic (DA) technique could be applied to eliminate multiplication at the expense of addition.

7.2.4 More Appropriate Metrics for ECG Classification Performance Analysis

Throughout Chapter 4 and 5, three main metrics of classification performance are primarily used - accuracy (Chapter 4 and 5), specificity and sensitivity (Chapter 5). It

may seem sufficient when applying these three metrics in general machine learning applications. But the choice of metric should depend on the applications. Recall that our engineering solution is to produce alarm when abnormality is detected. To its extreme, such abnormality in ECGs may cause fatal sudden death of the patient. It is extremely critical to assure such solution produces minimum number of FP and FN. Compared to FP which would not cost patient's life but consumable resources, FN should be avoided at all time at best effort. Despite so, in the thesis accuracy simply assumes equal loss between FP and FN, thus no actual concern was made to take the severity of FN into account. Therefore, in our future work more appropriate metrics like FP and FN should be considered to evaluate the ECG classification performance instead of just accuracy, specificity and sensitivity for more realistic analysis. Note that an integrated approach based on machine learning methods to reduce the rate of false-positive alarm to clinically useful levels is discussed in [6], which may help our future work.

7.2.5 More Robust Evaluation Method in Finding Optimal Set of Spectral Energy

In Section 4.3, four different approaches of deriving spectral energy of an entire ECG complex are covered. To evaluate the optimal feature set for each approach, best classification accuracy was used to perform the assessment. It is an easy-to-use metric to pinpoint which feature set provides the highest accuracy, and therefore this optimal one may then be selected for associated lead scenario. However, it may not be the best strategy to search for the optimal feature set. To achieve more robust evaluation of the experiments, Interquartile Ranges (IQRs) (especially, its median) of the distributions of accuracies should be performed to obtain reasonable representative of the distribution. Doing so avoids the dependence on the selection of extrema (in this case, maximum accuracy) from sets of experiments, and further avoids chance findings from multiple testing. More well-generalisable features and leads for robust use in clinical practice can then be possible.

7.2.6 A Better Statistic Model Exploration for Spectral Energy Variation

In Section 4.4.2, statistically the bias of spectral energy from ground truth under misdetection was assumed to follow Gaussian distribution. This is given the conditions where, according to Central Limit Theorem, observation (the bias) does not depend on the values of the other observations, because it comes from a separate/individual patient to the rest. However, in reality it may not be practical. Improvement can be made by exploiting more realistic distribution model under certain conditions after observing thoroughly how the distribution of the data look like (e.g. histogram). More precise and realistic conclusion for spectral energy variation under misdetection can then be made.

Appendix A

ECG Heartbeat Segmentation

A separately developed heartbeat segmentation algorithm is introduced here. It is a technique that localise the heartbeats (PQRST complex) in a data sequence of heart signal. Given a template of a heartbeat and the entire data sequence of the heart signal, the algorithm outputs a set of boundaries that indicate the on and off of the consecutive heartbeats. Following lists the pseudocode of the algorithm.

```
function complexOnOff = HeartbeatSegmentation(ecgData, templateTimeVec)
% Input:  ecgData           data vector of ECG data sequence
%         templateTimeVec   time vector of ECG template for segmentation
% Output: complexOnOff      time instant of the on and off of heartbeats

% ## Step 1 ## Initial variables
templateInterval = 50;      % subject to change in different scenarios
templateLength = length(templateTimeVec);
ecgDataLength = length(ecgData);

% ## Step 2 ## Set up a template that is chosen by the user
ecgTemplate = ecgData(templateTimeVec);

% ## Step 3 ## Initial Heartbeat Synchronisation
for i = 1 : templateInterval : templateLength
    % extract i to (i+templateLength-1) from ecgData as tempSample,
    % calculate the difference between tempSample and ecgTemplate, and sum them up
end
% Among all sums, find out the minimum and thus able to locate the first heartbeat
% Therefore, we have
complexOnOff(1) = onset of the first heartbeat;
complexOnOff(2) = complexOnOff(1) + templateLength;

% ## Step 4 ## Run TDMG on the first heartbeat to get the time instant of its R peak,
% and set up a increment for future use.
% Therefore, we have
timeRPeak = time instant of 1st R peak;
increment = complexOnOff(2) - timeRPeak;

% ## Step 5 ## Looping to extract the on and off time instants.
indexOnOff = 3;             % index of the boundary of heartbeats
while(1)
```

```
33      complexOnOff(indexOnOff) = complexOnOff(indexOnOff-1) + templateLength;
34      if complexOnOff(indexOnOff) > ecgDataLength
35          break;
36      end
37      % Run TDMG to obtain the time instant of the current R peak
38      timeRPeakCur = time instant of current R peak
39
40      % Correct complexOnOff according to the position of current R peak
41      complexOnOff(indexOnOff) = timeRPeakCur + increment;
42
43      indexOnOff = indexOnOff + 1;
44  end
```

Appendix B

Lead Arrangement for Four Sets of Experiments of Spectral Energy Derivation

In Section 4.3, participating leads in each lead scenario in each set of experiment are selected after exhaustive simulation. Specifically, depending on the experiment and classifier, the choice of participating leads varies from one scenario to another, ultimately up to 5 lead scenario. As presented in Table B.1 and Table B.2, the selected leads are listed in order with respect to lead scenario for our classifiers, covering all 4 sets of experiments.

Table B.1: Lead Arrangement for Four Sets of Experiments of Spectral Energy Derivation.

Lead	**Set 1**				
Scen.	**LDA**	**QDA**	**SVM**$_L$	**SVM**$_Q$	**k-NN**
1	aVL	V1	aVL	aVL	V1
2	III, aVF	III, aVL	III, aVF	III, aVF	II, V1
3	I, aVR, V1	III, aVL, V1	III, aVL, V1	aVL, aVF, V1	I, II, V5
4	I, II, III, V1	I, V1, V3, V5	I, II, aVR, V1	I, aVR, V1, V6	I, II, aVR, V5
5	I, II, III, aVL, V1	I, III, aVL, aVF, V1	I, III, aVR, V1, V2	I, III, aVL, aVF, V5	I, II, aVR, V1, V5
Lead	**Set 2 (Case 1)**				
Scen.	**LDA**	**QDA**	**SVM**$_L$	**SVM**$_Q$	**k-NN**
1	aVR	aVR	II	aVR	V5
2	I, aVR	I, aVR	II, aVL	aVR, aVL	I, V5
3	aVR, aVL, V1	I, aVR, V4	II, aVF, V1	I, V5, V6	I, aVR, V5
4	I, aVR, aVL, V1	III, aVL, aVF, V4	I, II, III, V1	I, aVR, aVL, V5	I, aVR, V5, V6
5	I, III, aVR, V1, V4	II, aVR, aVF, V4, V5	I, II, aVL, V1, V3	I, aVR, aVL, V5, V6	aVR, aVL, V1, V5, V6
Lead	**Set 2 (Case 2)**				
Scen.	**LDA**	**QDA**	**SVM**$_L$	**SVM**$_Q$	**k-NN**
1	aVR	aVR	II	aVR	II
2	II, aVL	aVR, aVL	I, aVR	aVR, aVL	I, V5
3	aVR, aVL, V1	I, aVR, V4	II, aVF, V1	I, V5, V6	I, aVR, V5
4	II, III, V1, V2	III, aVL, aVF, V4	I, II, III V1	I, aVR, aVL, V5	I, aVR, V5, V6
5	I, III, aVR, V1, V4	II, aVR, aVF, V4, V5	I, II, aVL, V1, V3	I, aVR, aVL, V5, V6	I, aVR, aVL, V5, V6

Table B.2: Lead Arrangement for Four Sets of Experiments of Spectral Energy Derivation. (Contin.)

Lead	**Set 3**				
Scen.	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
1	aVR	aVR	aVR	aVR	V5
2	II, aVL	aVR, aVL	III, aVF	aVL, V5	I, V5
3	II, III, V1	II, aVL, V4	II, III, V1	I , V5, V6	aVR, V5, V6
4	I , aVR, aVL, V1	III, aVF, V3, V4	II, III, V1, V3	aVL, V2, V5, V6	I , aVR, V5, V6
5	I , aVR, aVL, V1, V3	I , II, III, aVL, V2	II, III, V1, V2, V3	I , aVR, V2, V5, V6	aVR, aVL, V1, V5, V6
Lead	**Set 4**				
Scen.	**LDA**	**QDA**	**SVM$_L$**	**SVM$_Q$**	**k-NN**
1	III	aVR	II	aVR	aVR
2	II, III	III, aVR	II, III	aVR, V2	aVR, aVL
3	II, III, V1	III, aVR, aVL	II, III, V2	aVR, aVL, V2	aVR, aVL, V3
4	II, III, V1, V2	III, aVR, aVL, V4	II, III, aVL, V2	III, aVR, aVL, V2	aVR, aVL, V2, V3
5	I , II, III, V1, V2	III, aVR, aVL, aVF, V4	II, III, aVL, aVF, V2	II, III, aVR, aVL, V2	III, aVR, aVL, V2, V3

Appendix C

Experimental Results for Statistical Analysis of the Variation of Spectral Energy under Misdetection

Detailed results that are not covered in Section 4.4.2 regarding the figures of bias distribution, mean bias, 95% confidence interval of each lead and each case for different spectral energy of wave components of interest are illustrated in this appendix. Figure C.1 to C.7 are regarded as complementary materials to Section 4.4.2.

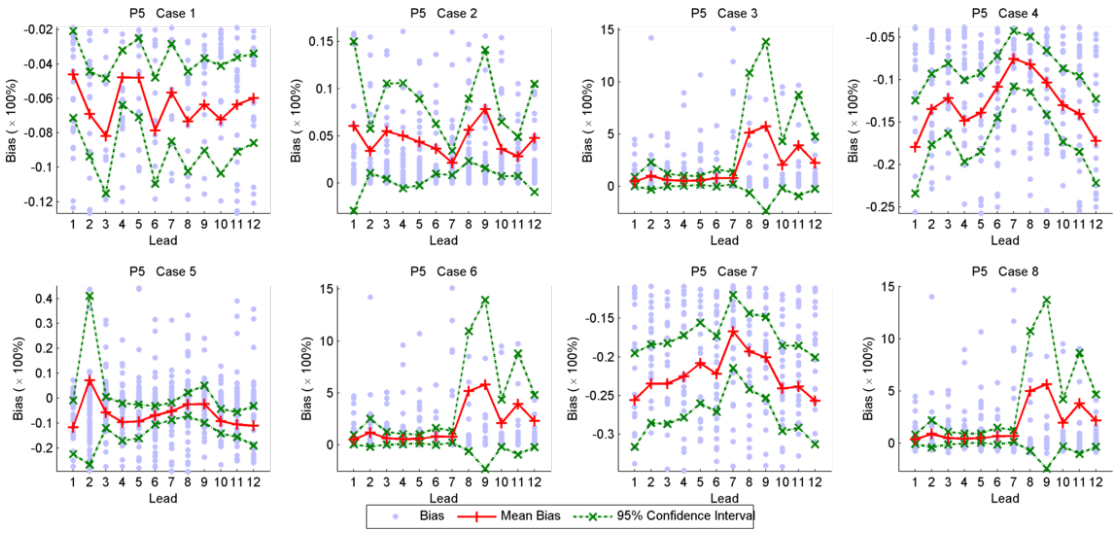


Figure C.1: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of P_5.

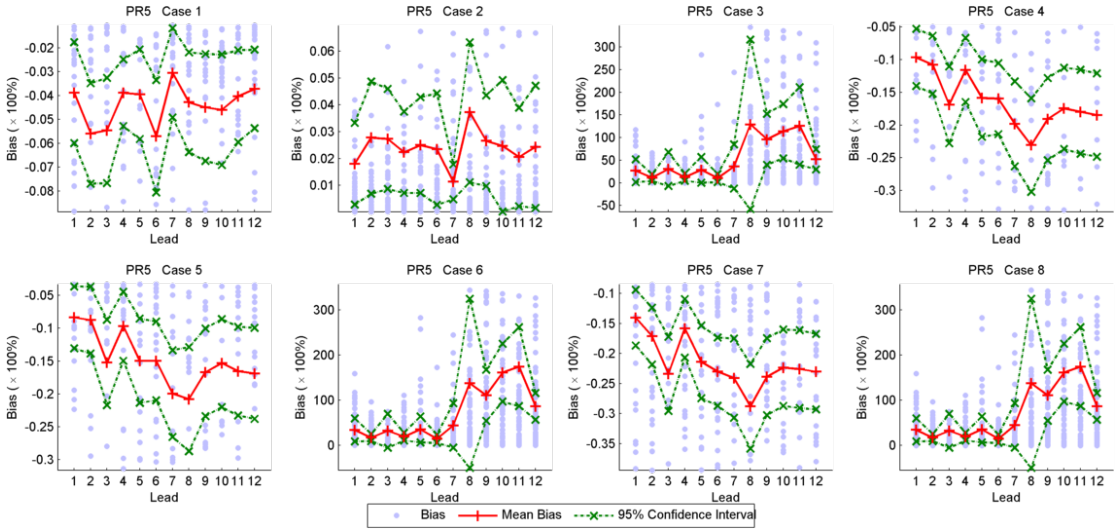


Figure C.2: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of PR_5.

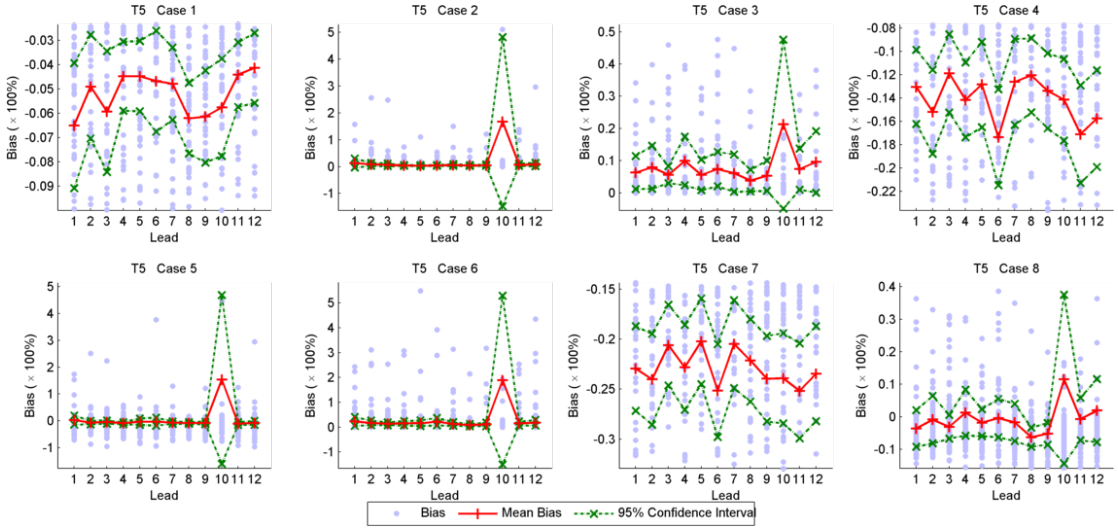


Figure C.3: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of T_5.

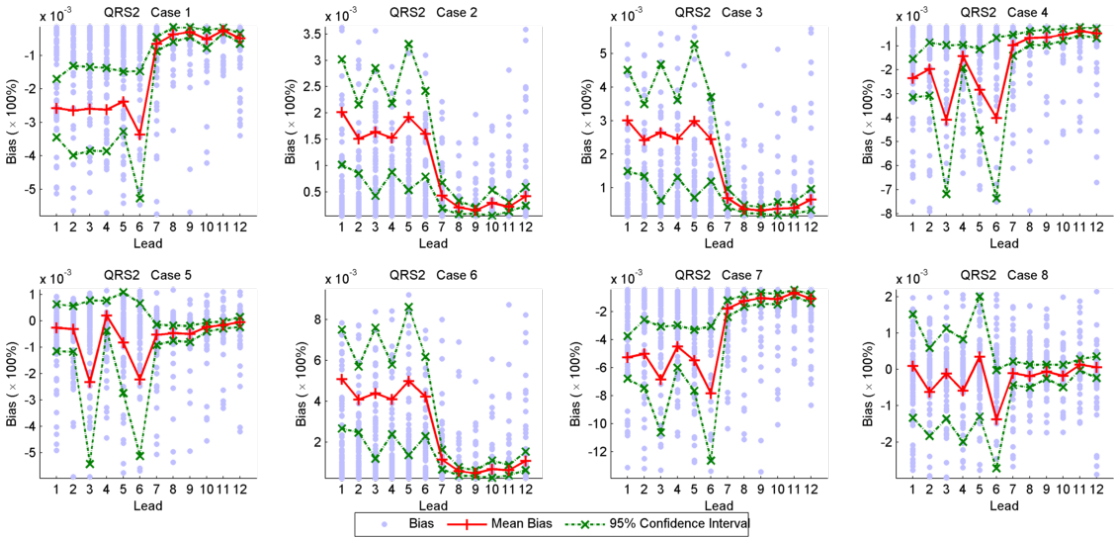


Figure C.4: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of QRS_2.

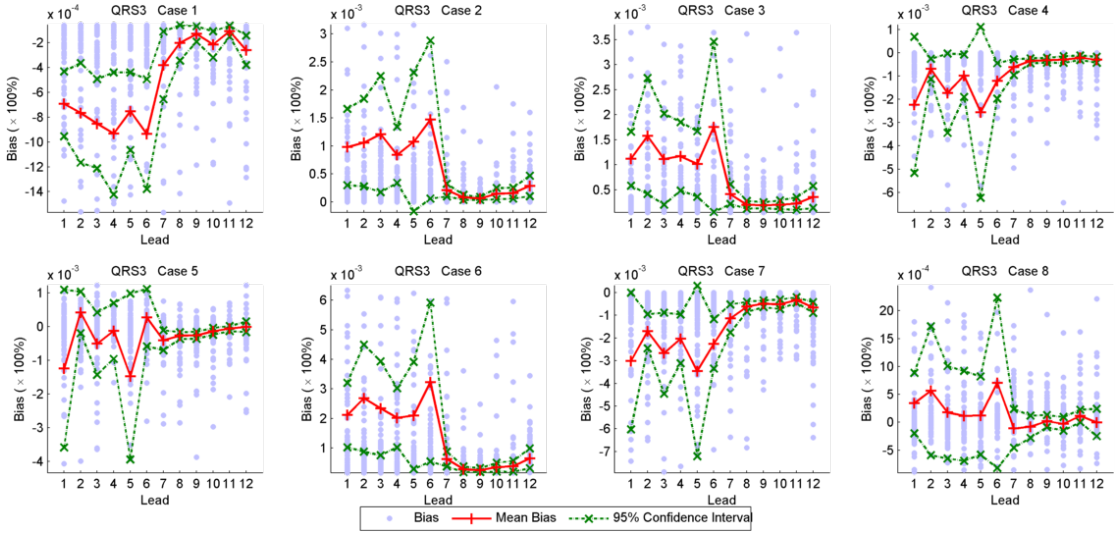


Figure C.5: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of QRS_3.

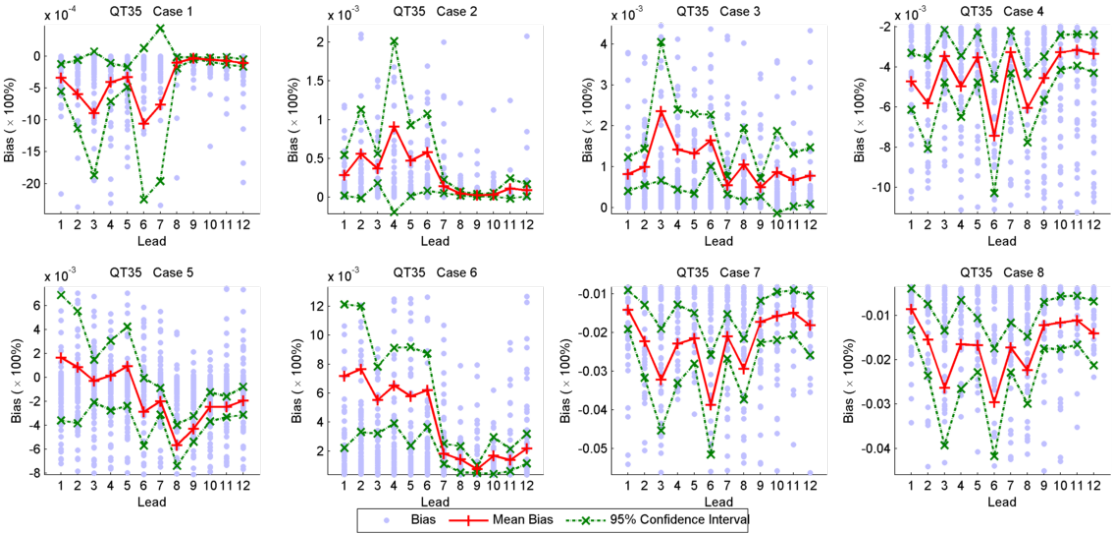


Figure C.6: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of QT_{35}.

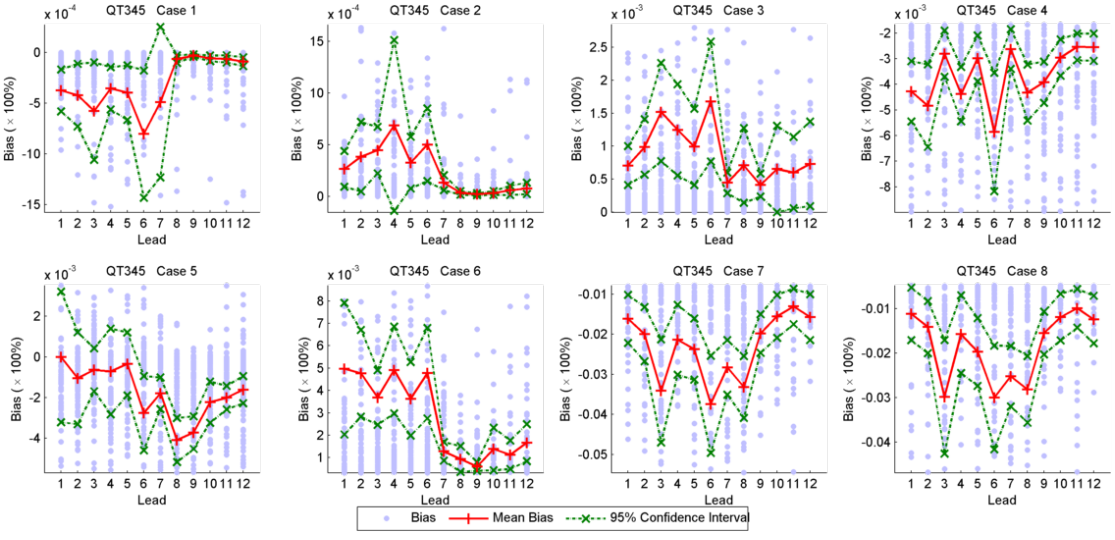


Figure C.7: Bias distribution of each lead as well as the associated mean bias and 95% confidence interval of eight cases of QT_{345}.

Appendix D

Experimental Results for Spectral Energy-based Classification Augmented with Feature Groups via Feature Selection Algorithms

Detailed results that are not covered in Section 5.4.3.2 are described and explained in this appendix. According to the experimental procedure in Section 5.4.3.1, ReliefF and InfoGain effectively output a list of weighted features. To examine the classification performance, all features are taken into account in descending order with respect to the weights. But since the entire process involve significantly huge number of assessments, we opt to pick up the feature subset that achieves the highest accuracy within each weighted feature unit (Figure 5.4). On the other hand, because CFS and FCBF directly output the optimal feature subset within each feature unit, no further action is needed. Subsequently, we obtain the highest accuracies for each feature group, each lead scenario and each classifier, which are shown in the following figures.

To give better insight into the results, following discussion is made.

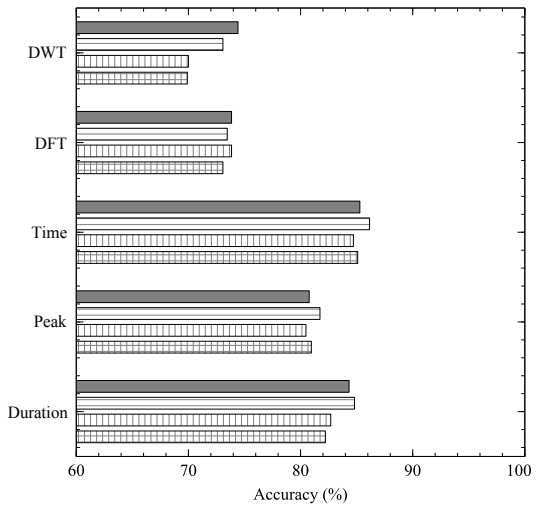
- *LDA*: Overall, 1 lead scenario is shown to exhibit fairly low accuracy. That could be due to the fact that features extracted from only one lead may not facilitate classification, regardless of the feature group. For the rest of the lead scenarios, accuracy tends to increase slightly along with the lead scenario on average. In 5 lead scenario, the highest accuracy can be observed under Peak and CFS.
- *QDA*: It is interesting to observe that, DWT and DFT as feature groups do not perform well compared to the other groups under all lead scenarios. Time seems to outperform the rest of the groups at all times.

- SVM_L: Similar to LDA, 1 lead scenario seems to present the lowest accuracy among all. Accuracy grows gradually along with lead scenario. The highest accuracy can be observed under DFT and ReliefF in 4 and 5 lead scenario.

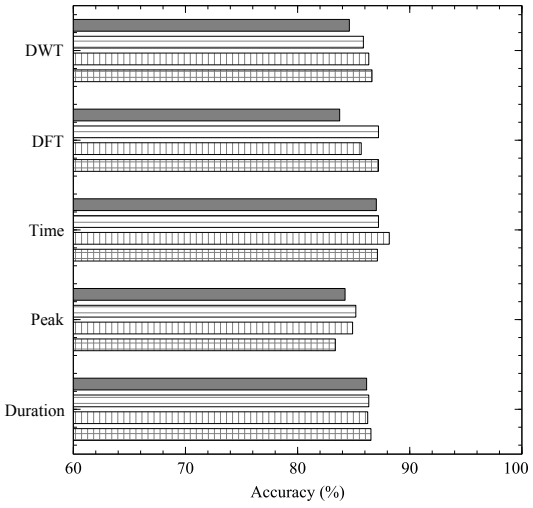
- SVM_Q: For SVM_Q, 1 lead scenario performs slightly worse than other lead scenarios. Accuracy in other lead scenarios improves from smaller to bigger lead scenario. Note that, interestingly 3 lead scenario performs not as expected.

- *k-NN*: As other classifiers, overall 1 lead scenario exhibits the lower accuracies. However, accuracies do not improve along with lead scenario. It can be seen that only 2 lead scenario and 5 lead scenario perform as expected; 3 and 4 lead scenario, on the other hand, show decreasing performance.

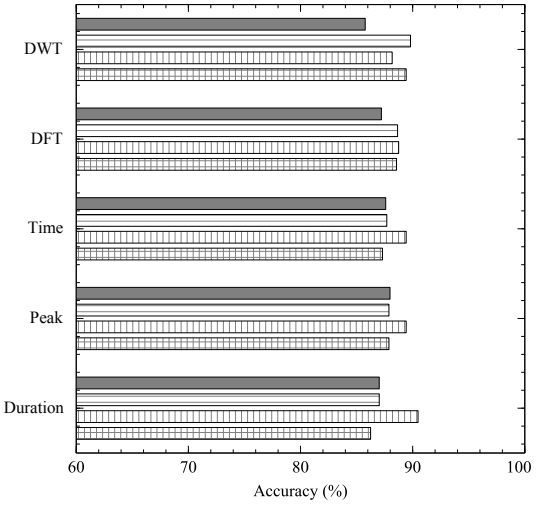
To conclude, classification accuracy depends heavily on classifier and lead scenario. Also, it seems that differences in performance between feature selection algorithms do not exist in general, except situations regarding specific feature group, for instance ReliefF and InfoGain for DFT under 3 and 5 lead scenario of LDA. Last but not least, with all those results, we manage to carry out a higher extraction of the results and present them in Table 5.7. That means, only the highest accuracies in each lead scenario from Figure D.1 to Figure D.5 are picked up and specifically listed, in order for us to draw further conclusion in Section 5.4.3.2.



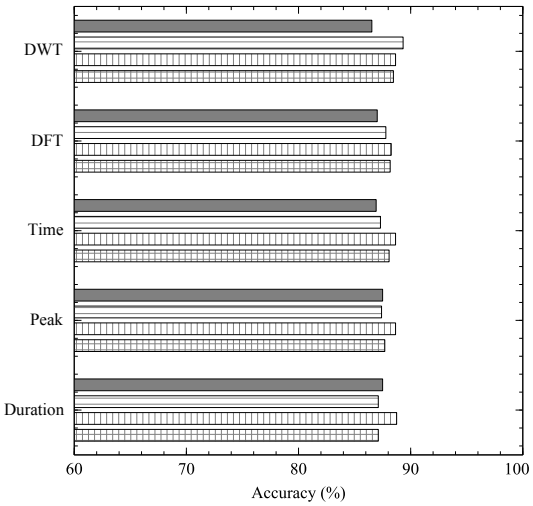
(a) 1 Lead Scenario



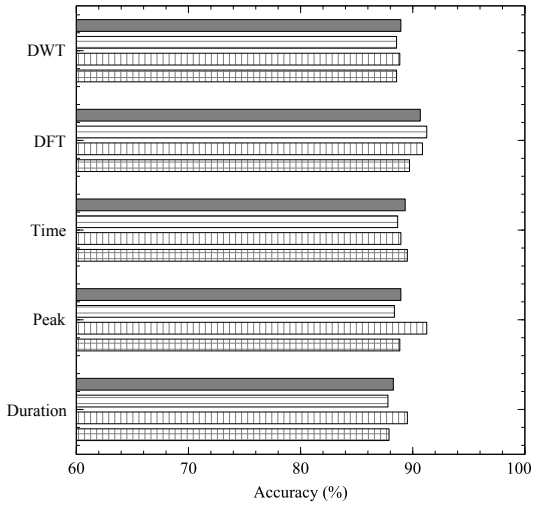
(b) 2 Lead Scenario



(c) 3 Lead Scenario



(d) 4 Lead Scenario



(e) 5 Lead Scenario

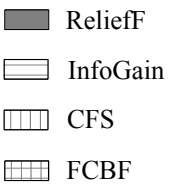
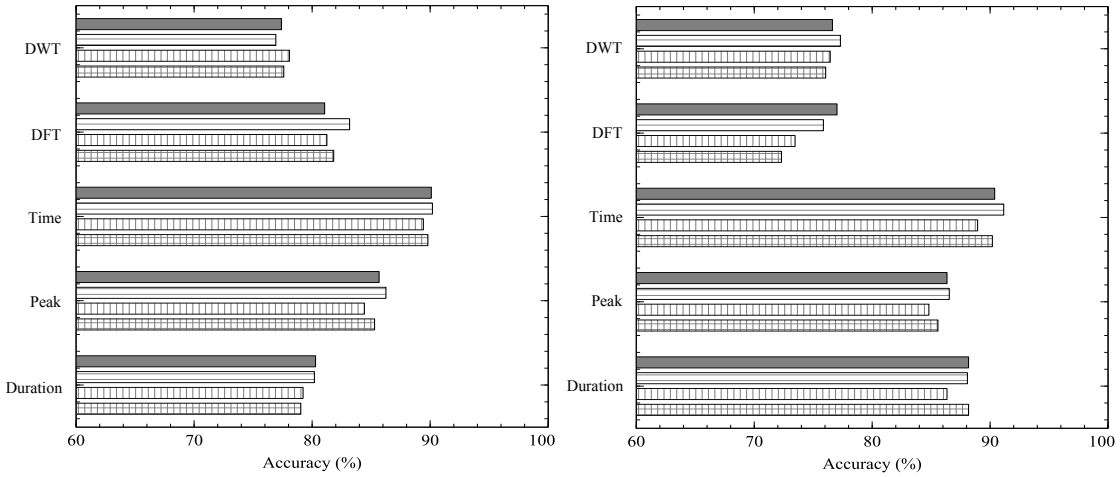
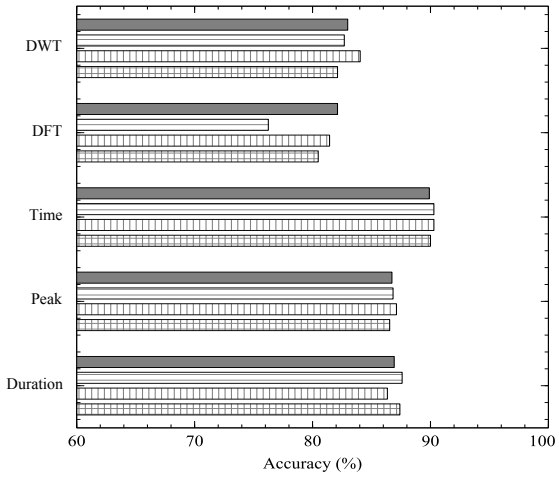


Figure D.1: The highest accuracies obtained with features from five feature groups under five lead scenarios of LDA after feature selection.

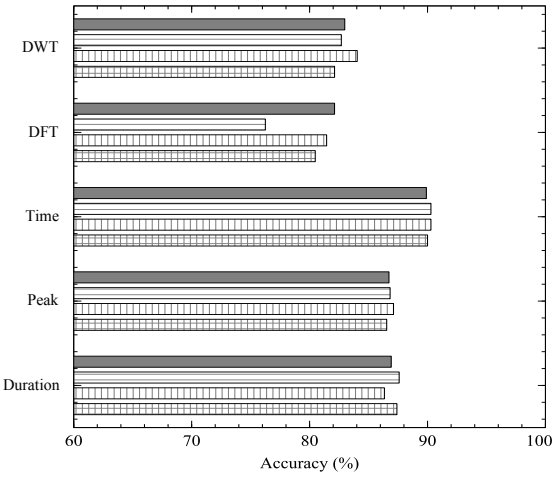


(a) 1 Lead Scenario

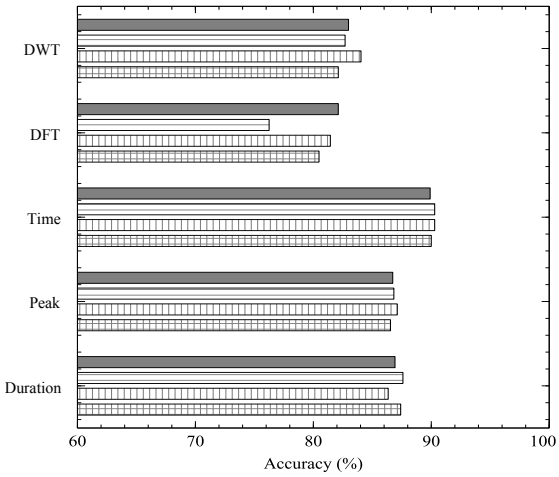
(b) 2 Lead Scenario

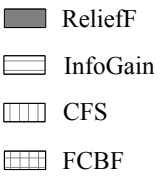


(c) 3 Lead Scenario



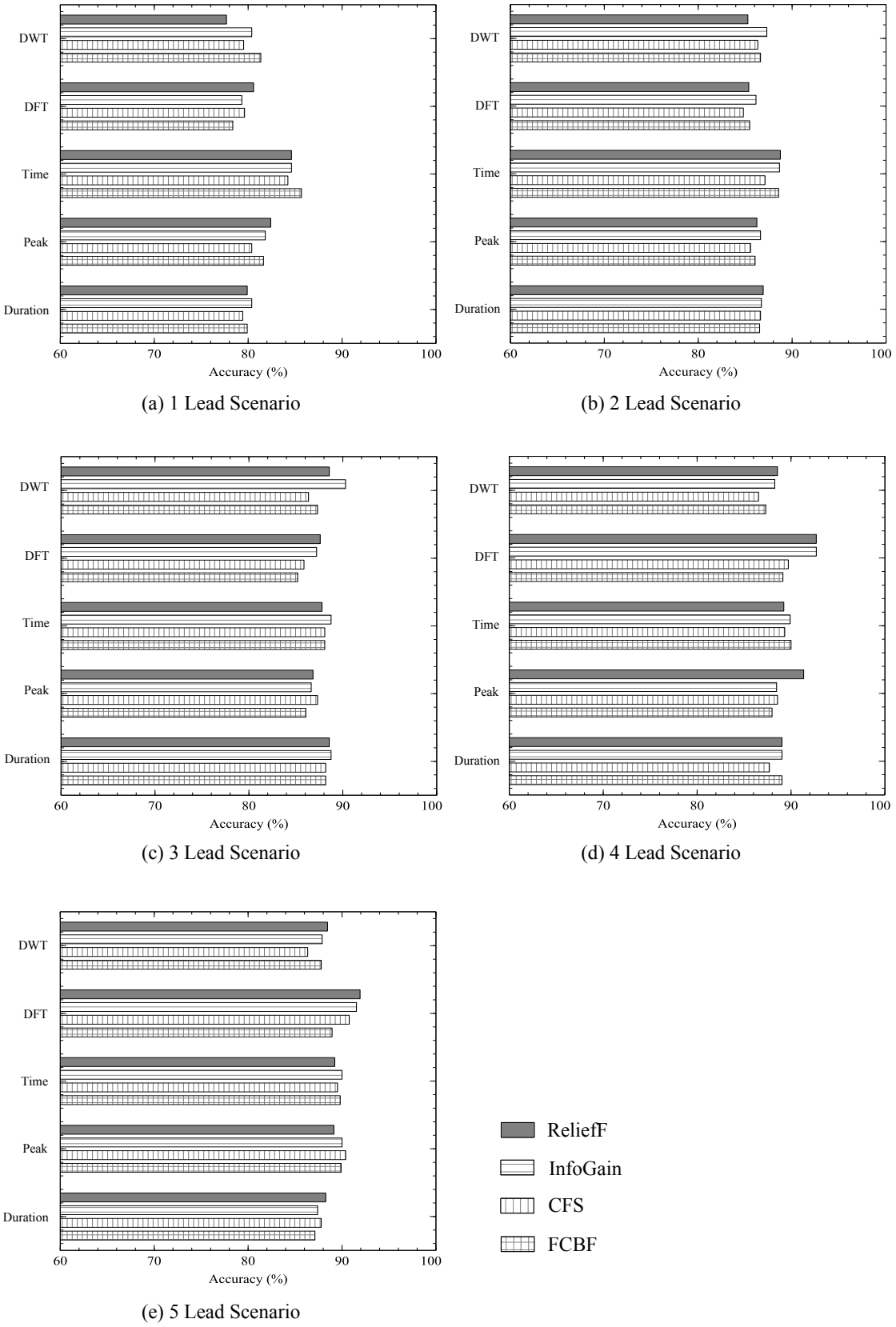
(d) 4 Lead Scenario





(e) 5 Lead Scenario

Figure D.2: The highest accuracies obtained with features from five feature groups under five lead scenarios of QDA after feature selection.



(a) 1 Lead Scenario

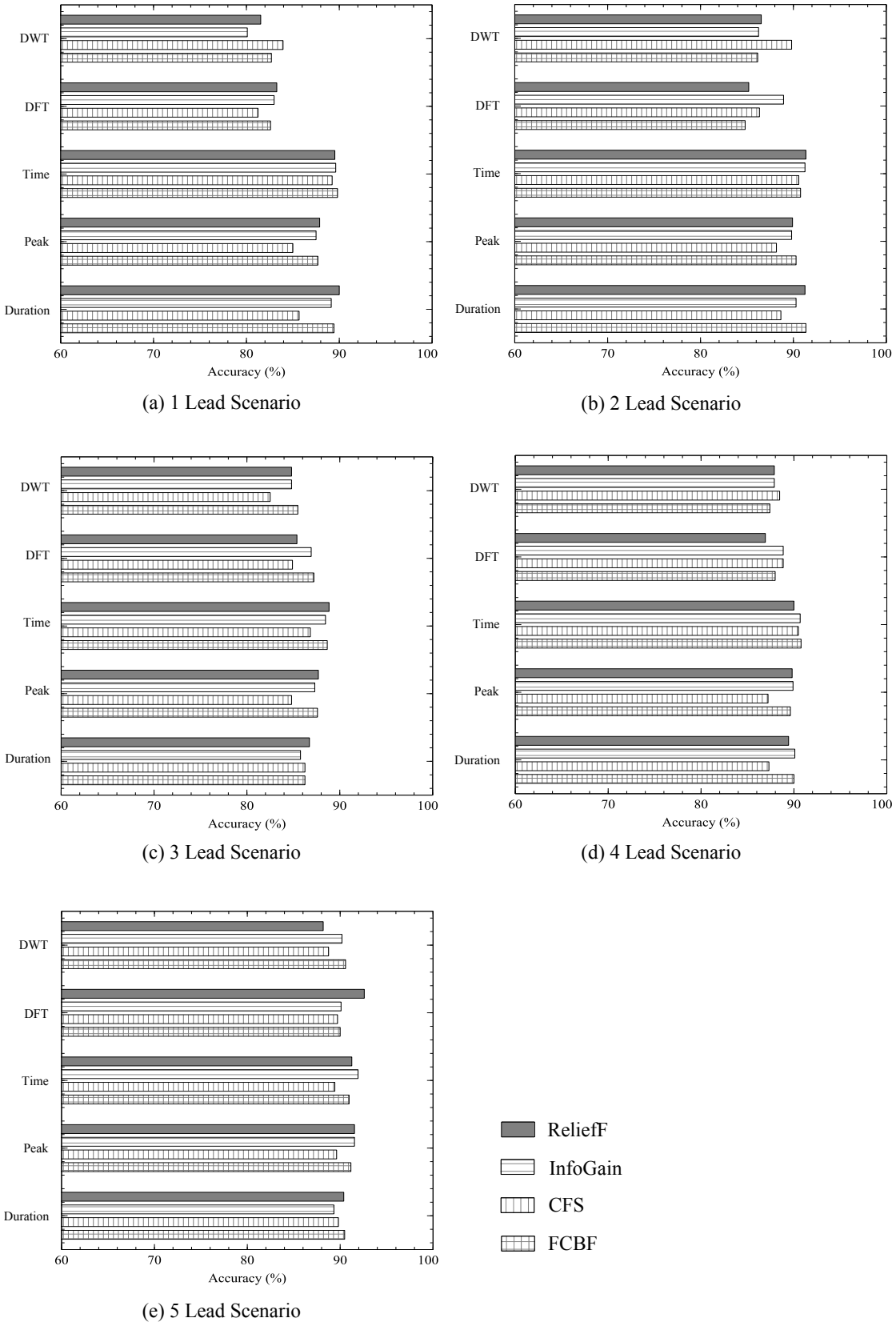
(b) 2 Lead Scenario

(c) 3 Lead Scenario

(d) 4 Lead Scenario

(e) 5 Lead Scenario

Figure D.3: The highest accuracies obtained with features from five feature groups under five lead scenarios of SVM_L after feature selection.



(a) 1 Lead Scenario

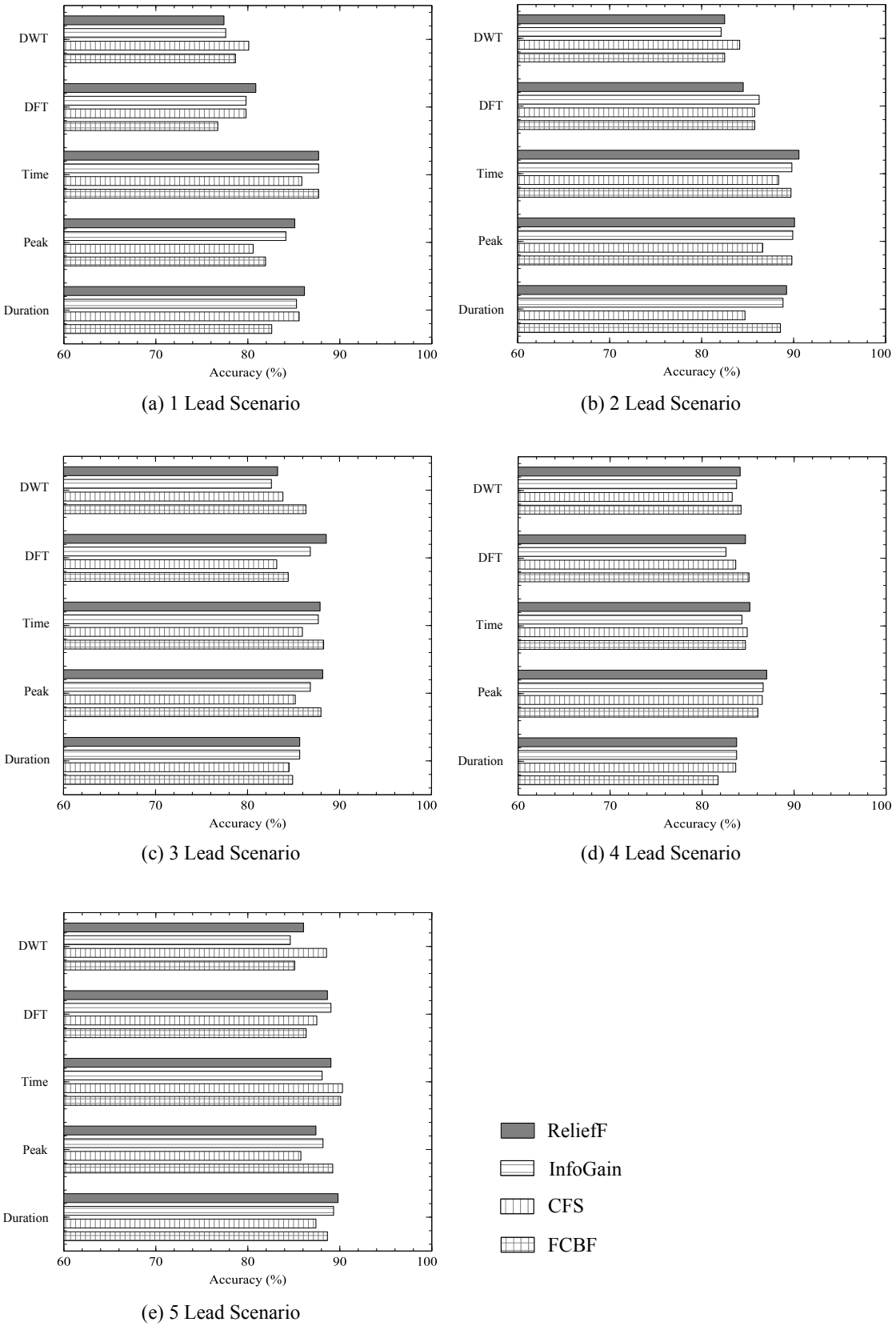
(b) 2 Lead Scenario

(c) 3 Lead Scenario

(d) 4 Lead Scenario

(e) 5 Lead Scenario

Figure D.4: The highest accuracies obtained with features from five feature groups under five lead scenarios of SVM_Q after feature selection.



(a) 1 Lead Scenario

(b) 2 Lead Scenario

(c) 3 Lead Scenario

(d) 4 Lead Scenario

(e) 5 Lead Scenario

Figure D.5: The highest accuracies obtained with features from five feature groups under five lead scenarios of k-NN after feature selection.

References

[1] K. Maharatna, E. Mazomenos, J. Morgan and S. Bonfiglio. Towards the development of next-generation remote healthcare system: some practical considerations. In *Proc. IEEE Int'l Symposium on Circuits and Systems (ISCAS 2012)*, pages 1–4, 2012.

[2] R. Sokullu, M. Akkas, and H. Cetin. Wireless patient monitoring system. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on*, pages 179–184, 2010.

[3] L. Hanh, M. Kuttel and G. Chandran. An electronic health care - cardiac monitoring system. In *Communications Workshops (ICC), 2010 IEEE International Conference on*, pages 1–5, 2010.

[4] S. Dagtas, G. Pekhteryev, Z. Sahinoglu, H. Cam, and N. Challa. Real-time and secure wireless health monitoring. *International Journal of Telemedicine Application*, pages 1–10, 2008.

[5] N. Vrcek, M. Velic, and Z. Stapic. Integrated mobile electrocardiography. In *Proceedings of the 30th MIPRO International Convention on Computers in Technical Systems*, 2007.

[6] D. Clifton, D. Wong, L. Clifton, S. Wilson, R. Way, R. Pullinger and L. Tarassenko. A large-scale clinical validation of an integrated monitoring system in the Emergency Department. *IEEE Journal of Biomedical and Health Informatics*, 17:835–842, 2013.

[7] L. Clifton, D. Clifton, M. Pimentel, P. Watkinson and L. Tarassenko. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. *IEEE Journal of Biomedical and Health Informatics*, 18:722–730, 2014.

[8] S. McLean, D. Protti and A. Sheikh. Telehealthcare for long term conditions [Online]. Available at: http://www.bmj.com/content/342/bmj.d120, [Accessed: Jan. 2015].

[9] G. Clifford and D. Clifton. Wireless technology in disease management and medicine. *Annual Review of Medicine*, 63:479–492, 2012.

[10] WHO. Cardiovascular disease [Online]. Available at: http://www.who.int/cardiovascular_diseases/en/, [Accessed: Sep. 2012].

[11] W. Rosamond, K. Flegal, G. Friday, K. Furie, A. Go, K. Greenlund, N. Haase, M. Ho, V. Howard, B. Kissela, B. Kissela, S. Kittner, D. Lloyd-Jones, M. McDermott, J. Meigs, C. Moy, G. Nichol, C. J. ODonnell, V. Roger, J. Rumsfeld, P. Sorlie, J. Steinberger, T. Thom, S. Wasserthiel-Smoller and Y. Hong. Heart disease and stroke statistics-2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*, 115:69–171, 2007.

[12] WHO. *Prevention of Cardiovascular Disease - Pocket Guidelines for Assessment and Management of Cardiovascular Risk*. WHO Press, Geneva, 2007.

[13] WHO. Cardiovascular disease - strategic priorities [Online]. Available at: http://www.who.int/cardiovascular_diseases/priorities/en/, [Accessed: Nov. 2012].

[14] J. Leal, R. Luengo-Fernandez, A. Gray, S. Petersen, M. Rayner. Economic burden of cardiovascular diseases in the enlarged European Union. *European Heart Journal*, 27:1610–1619, 2006.

[15] AHA. Heart disease and stroke statistics - economic cost of cardiovascular diseases [Online]. Available at: https://my.clevelandclinic.org/Documents/heart, [Accessed: Sep. 2012].

[16] H. Kennedy. The evolution of ambulatory ECG monitoring. *Progress in Cardiovascular Diseases*, 56:127–132, 2013.

[17] CHIRON. CHIRON: Cyclic and persona-centric Health management: Integrated appRoach for hOme, mobile and clinical eNvironments [Online]. Available at: http://www.chiron-project.eu, [Accessed: Mar. 2014].

[18] N. Saranummi. In the spotlight: health information systems. *IEEE Reviews in Biomedical Engineering*, 4:17–19, 2011.

[19] M. Hanson, H. Powell, A. Barth, K. Ringgenberg, B. Calhoun, J. Aylor and J. Lach. Body area sensor networks: challenges and opportunities. *Computer*, 42:58–65, 2009.

[20] N. Selvaraj, A. Jaryal, J. Santhosh, K. Deepak and S. Anand. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering & Technology*, 32:479–84, 2008.

[21] G. Lu, F. Yang, J. Taylor and J. Stein. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of Medical Engineering & Technology*, 33:634–41, 2009.

[22] T. Chen, E. Mazomenos, K. Maharatna, S. Dasmahapatra and M. Niranjan. Design of a low-power on-body ECG classifier for remote cardiovascular monitoring systems. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 3:75–85, March 2013.

[23] Z. Loring, S. Chelliah, R. Selvester, G. Wagner and D. Strauss. A detailed guide for quantification of myocardial scar with the Selvester QRS score in the presence of electrocardiogram confounders. *Journal of Electrocardiology*, 44:544–554, 2011.

[24] L. Tarassenko, D. Clifton, P. Bannister, S, King and D. King. Novelty Detection. *Encyclopedia of Structural Health Monitoring*, pages 1–23, 2009.

[25] Roger, V. L. and et al. Heart disease and stroke statistics–2012 update: a report from the American Heart Association. *Circulation*, 125:e2–e220, 2012.

[26] V. Bono, E. B. Mazomenos, T. Chen, J. A. Rosengarten, A. Acharyya, K. Maharatna, J. M. Morgan, and N Curzen. Development of an automated updated Selvester QRS scoring system using SWT-Based QRS fractionation detection and classification. *IEEE Journal of Biomedical Health Informatics*, 18:193–204, 2014.

[27] G. Yang. *Body Sensor Networks*. Springer, London, 2006.

[28] R. Chakravorty. A programmable service architecture for mobile medical care. In *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW'06)*, pages 532–536, March 2006.

[29] F. Wang, L. Docherty, K. Turner, M. Kolberg and E. Magill. Services and policies for care at home. In *Pervasive Health Conference and Workshops*, pages 1–10, Nov 2006.

[30] P. Kuryloski, A. Giani, R. Giannantonio, K. Gilani, R. Gravina, V. Seppa, E. Seto, V. Shia, C. Wang, P. Yan, A. Yang, J. Hyttinen, S. Sastry, S. Wicker and R. Bajcsy. DexterNet: an open platform for heterogeneous body sensor networks and its applications. In *Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, pages 92–97, June 2009.

[31] Oxleas NHS UK. Long term conditions [Online]. Available at: http://www.oxleas.nhs.uk/long-term-conditions/, [Accessed: Jan. 2015].

[32] B. Drew, R. Califf, M. Funk, E. Kaufman, M. Krucoff, M. Laks, P. Macfarlane, C. Sommargren, S. Swiryn and G. Van Hare. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. *Circulation*, 110:2721–2746, 2004.

[33] A. Auricchio, and et al. Long-term clinical effect of hemodynamically optimized cardiac resynchronization therapy in patients with heart failure and ventricular conduction delay. *Journal of the American College of Cardiology*, 39:2026–2033, 2002.

[34] B. Strauer, Y. Muhammad , M. Christiana. The acute and long-term effects of intracoronary stem cell transplantation in 191 patients with chronic heARt failure: the STAR-heart study. *European Journal of Heart Failure*, 12:721–729, 2010.

[35] A. Forkan, I. Khalil and Z. Tari. Context-aware cardiac monitoring for early detection of heart diseases. In *Computing in Cardiology Conference (CinC), 2013*, pages 277–280, 2013.

[36] W. Liang, S. Hu, Z. Shao and J. Tan. A real-time cardiac arrhythmia classification system with wearable electrocardiogram. In *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2011 IEEE International Conference on*, pages 102–106, 2011.

[37] M. Rosu and S. Pasca. A WBAN-ECG approach for real-time long-term monitoring. In *Advanced Topics in Electrical Engineering (ATEE), 2013 8th International Symposium on*, pages 1–6, 2013.

[38] P. Bella and et al. Management of ventricular tachycardia in the setting of a dedicated unit for the treatment of complex ventricular arrhythmias: long term outcome after ablation. *Circulation*, pages 1–26, 2013.

[39] P. Zimetbaum and A. Goldman. Ambulatory arrhythmia monitoring: choosing the right device. *Circulation*, 122:1629–1636, 2010.

[40] M. Crawford, et al. ACC/AHA Guidelines for Ambulatory Electrocardiography: Executive Summary and Recommendations : A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the Guidelines for Ambulatory Electrocardiography) Developed in Collaboration With the North American Society for Pacing and Electrophysiology. *Circulation*, 100:886–893, 1999.

[41] R. Fensli, E. Gunnarson and T. Gundersen. A wearable ECG-recording system for continuous arrhythmia monitoring in a wireless tele-home-care situation. In *Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on*, pages 407–412, 2005.

[42] S. Jeong and et al. An integrated healthcare system for personalized chronic disease care in home-hospital environments. *IEEE Transactions on Information Technology in Biomedicine*, 16:572–585, 2012.

[43] FDA. Guidance for the content of premarket submissions for software contained in medical devices [Online]. Available at: http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm089543.htm, [Accessed: Mar. 2015].

[44] FDA. General principles of software validation: final guidance for industry and FDA staff [Online]. Available at: http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm085281.htm, [Accessed: Jan. 2015].

[45] M. Jaakko and P. Robert. *Bioelectromagnetism*. Oxford University Press, New York, 1995. [Online] http://www.bem.fi/edu/book.htm.

[46] B. Tomas and E. Neil. *12-Lead ECG The Art of Interpretation*. Jones and Bartlett Learning, London, 2001.

[47] P. Macfarlane, A. Van Oosterom, P. Olle, J. Michiel and C. John. *Comprehensive Electrocardiology*. Springer, London, 2 edition, 2011.

[48] F. Rumulo. *Basic and Bedside Electrocardiography*. Wolters Kluwer Lippincott Williams and Wilkins, London, 2009.

[49] L. Sörnmo and P. Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological applications*. Elsevier Academic Press, London, 2005.

[50] E. Pueyo, R. Bailón, and E. Gil. Signal processing guided by physiology: making the most of cardiorespiratory signals. *IEEE Signal Processing Magazine*, pages 136–142, 2013.

[51] S. Greenwald. Development and analysis of a ventricular fibrillation detector. *M.S. thesis, MIT Dept. of Electrical Engineering and Computer Science*, 1986.

[52] PhysioNet. Sudden cardiac death Holter database [Online]. Available at: http://www.physionet.org/cgi-bin/atm/ATM?database=sddb&tool=plot_waveforms, [Accessed: Jan. 2015].

[53] J. Venegas and R. Mark. HST.542J/2.792J/BE.371J/6.022J Quantitative physiology: organ transport systems, lecture notes from HST/MIT Open Courseware 2004 [Online]. Available at: http://ocw.mit.edu/courses/health-sciences-and-technology/, [Accessed: Oct. 2013].

[54] A. Waller. A demonstration on man of the electromotive changes accompanying the hearts beat. *The Journal of Physiology*, 8:229–234, 1887.

[55] W. Einthoven and K. Lint. Ueber das normale menschliche Elektrokardiogramm und ber die capillar-elektrometrische Untersuchung einiger Herzkranken. *Pflügers Archiv European Journal of Physiology* , 80:139–160, 1900.

[56] T. Lewis, J. Meakins and P. White. The excitatory process in the dog's heart. Part I. The auricles. *Philosophical Transactions of the Royal Society of London*, 205:375–420, 1914.

[57] T. Lewis and M. Rothschild. The excitatory process in the dog's heart. Part II. The ventricles. *Philosophical Transactions of the Royal Society of London*, 206:181–226, 1915.

[58] E. Goldberger. A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads. *American Heart Journal*, 23:483–492, 1942.

[59] D. Kreiseler R. Bousseljot and A. Schnabel. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1):317–318, 1995.

[60] D. Geselowitz. Dipole theory in electrocardiography. *The American Journal of Cardiology*, 14:301–306, 1964.

[61] R. Mason and I. Likar. A new system of multiple-lead exercise electrocardiography. *American Heart Journal*, 71:196–205, 1966.

[62] E. Frank. A new system of multiple-lead exercise electrocardiography. *Circulation*, 13:737–749, 1956.

[63] G. Dower. A lead synthesizer for the Frank system to simulate the standard 12-lead electrocardiogram. *Journal of Electrocardiology*, 1:101–116, 1968.

[64] N. Holter. New method for heart studies: continuous electrocardiography of active subjects over long periods is now practical. *Science*, 134:1214–1220, 1961.

[65] SCHILLER. SCHILLER's medilog® Holter system [Online]. Available at: http://www.schiller.ch/upload/medilog_EN.pdf, [Accessed: Mar. 2014].

[66] getemed. getemed CardioMem® CM 4000 [Online]. Available at: http://www.getemed.net/en/cardiology/cardiomemr-cm-4000/?R=1, [Accessed: Mar. 2014].

[67] PHILIPS. DigiTrak XT Holter recorder [Online]. Available at: http://www.healthcare.philips.com/gb_en/products/cardiography/products/holter/holter.wpd, [Accessed: Mar. 2014].

[68] G. Dower, A. Yakush, S. Nazzal and R. Jutzy. Deriving the 12-lead electrocardiogram from four (EASI) electrodes. *Journal of Electrocardiology*, 21:S182–S187, 1988.

[69] G. Wehr, R. Peters, K. Khalifé, A. Banning, V. Kuehlkamp, A. Rickards and U. Sechtem. A vector-based, 5-electrode, 12-lead monitoring ECG (EASI) is equivalent to conventional 12-lead ECG for diagnosis of acute coronary syndromes. *Journal of Electrocardiology*, 39:22–28, 2006.

[70] D. Feild, C. Feldman and B. Horácek. Improved EASI coefficients: their derivation, values, and performance. *Journal of Electrocardiology*, 35:23–33, 2002.

[71] A. Maybhate, S. Hao, S. Iwai, J. Lee, A. Guttigoli, K. Stein, B. Lerman and D. Christini. Detection of repolarization alternans with an implantable cardioverter defibrillator lead in a porcine model. *IEEE Transactions on Biomedical Engineering*, 52:1188–1194, 2005.

[72] P. Kowey and D. Kocovic. Cardiology patient pages. Ambulatory electrocardiographic recording. *Circulation*, 108:e1–3, 2003.

[73] A. Pantelopoulos and N. Bourbakis. A Survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40:1–12, 2010.

[74] D. Estrin and I. Sim. Open mHealth architecture: an engine for health care innovation. *Science*, 330:759–760, 2010.

[75] G. Mehl and A. Labrique. Prioritizing integrated mHealth strategies for universal health coverage. *Science*, 345:1284–1287, 2010.

[76] International Telecommunications Union. The world in 2010: ICT facts and figures [Online]. Available at: http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf, [Accessed: Feb. 2015].

[77] C. Meier, M. Fitzgerald and J. Smith. eHealth: extending, enhancing, and evolving health care. *Annual Review of Biomedical Engineering*, 15:359–382, 2013.

[78] World Health Organization. mHealth: new horizons for health through mobile technologies. *Global Observatory for eHealth Series*, 3, 2011.

[79] M. Tomlinson, M. Rotheram-Borus, L. Swartz and A. Alexander. Scaling up mHealth: where Is the evidence? *PLoS Medicine*, 10:1–5, 2013.

[80] M. Pavel and et al. The role of technology and engineering models in transforming healthcare. *IEEE Reviews in Biomedical Engineering*, 6:156–77, 2013.

[81] R. Lee and et al. A mobile care system with alert mechanism. *IEEE Transactions on Information Technology in Biomedicine*, 11:507–517, 2007.

[82] F. Hu, M. Jiang, M. Wagner and D. Dong. Privacy-preserving telecardiology sensor networks: toward a low-cost portable wireless hardware/software codesign. *IEEE Transactions on Information Technology in Biomedicine*, 11:619–627, 2007.

[83] Y. Huang and et al. Pervasive, secure access to a hierarchical sensor-based healthcare monitoring architecture in wireless heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, 27:400–411, 2009.

[84] A. Huang and et al. WE-CARE: an intelligent mobile telecardiology system to enable mHealth applications. *IEEE Journal of Biomedical and Health Informatics*, 18:693–702, 2014.

[85] A. Huang and et al. System light-loading technology for mHealth: manifold-learning based medical data cleansing and clinical trials in WE-CARE project. *IEEE Journal of Biomedical and Health Informatics*, 18:1581–1589, 2014.

[86] X. Liu and et al. A 457 nW near-threshold cognitive multi-functional ECG processor for long-term cardiac monitoring. *IEEE Journal of Solid-State Circuits*, 49:2422–2434, 2014.

[87] S. Hsu, Y. Ho, P. Chang and C. Su. Cardiac sensor SoC for mobile healthcare applications. *IEEE Journal of Solid-State Circuits*, 49:801–811, 2014.

[88] Y. Chen and et al. An injectable 64 nW ECG mixed-signal SoC in 65 nm for arrhythmia monitoring. *IEEE Journal of Solid-State Circuits*, 50:375–390, 2015.

[89] J. Semmlow. *Signals and Systems for Bioengineers - A Matlab-based Introduction.* Academic Press, London, 2012.

[90] M. Oppenheim and R. Schafer. *Discrete-Time Signal Processing.* Pearson, London, 3 edition, 2009.

[91] W. Cochran and et al. What is the fast Fourier transform? *Proceedings of the IEEE*, 55:1664–1674, 1967.

[92] H. Iwai and S. Ohmi. Silicon integrated circuit technology from past to future. *Microelectronics Reliability*, 42:465–491, 2002.

[93] Infiniti Research Limited. Global digital signal processor market 2012-2016 [Online]. Available at: http://www.researchandmarkets.com/reports/2510369/, [Accessed: Dec. 2013].

[94] G. Frantz. Digital signal processor trends. *Micro, IEEE*, 20:52–59, 2000.

[95] H. Ghasemzadeh, S. Ostadabbas, E. Guenterberg and A. Pantelopoulos. Wireless Medical Embedded Systems: A Review of Signal Processing Techniques for Classification. *IEEE Sensors Journal*, 13:423–437, 2013.

[96] S. Elder. Field Programmable Analog & Gallium Arsenide [Online]. Available at: http://www.planetanalog.com/author.asp?section_id=526&doc_id=559895, [Accessed: Dec. 2013].

[97] M. Unser and A. Aldroubi. A review of wavelets in biomedical applications. *Proceedings of the IEEE*, 84:626–638, 1996.

[98] M. Lakshmanan and H. Nikookar. A review of wavelets for digital wireless communication. *Wireless Personal Communications*, 37:387–420, 2006.

[99] A. Leung, F. Chau and J. Gao. A review on applications of wavelet transform techniques in chemical analysis: 1989-1997. *Chemometrics and Intelligent Laboratory Systems*, 43:165–184, 1998.

[100] T. Li, Q. Li, S. Zhu and M. Ogihara. A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, 4:49–68, 2002.

[101] D. Gabor. Theory of communication. Part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering*, 93:429–441, 1946.

[102] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[103] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

[104] R. Polikar. The story of wavelets. *Physics and Modern Topics in Mechanical and Electrical Engineering*, pages 192–197, 1999.

[105] R. Polikar. The engineer's ultimate guide to wavelet analysis: the wavelet tutorial [Online]. Available at: http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html, [Accessed: Dec. 2013].

[106] P. Addison. *The Illustrated Wavelet Transform Handbook*. IOP Publishing Ltd, Bristol, 2002.

[107] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2:50–61, 1995.

[108] O. Rioul. A discrete-time multiresolution theory. *IEEE Transactions on Signal Processing*, 41:2591–2606, 1993.

[109] P. Bentley and J. McDonnell. Wavelet transforms: an introduction. *Electronics & Communication Engineering Journal*, 6:175–186, 1994.

[110] S. Chen, H. Lee, C. Chen, C. Lin and C. Luo. A wireless body sensor network system for healthcare monitoring application. In *Biomedical Circuits and Systems Conference, IEEE*, pages 243–246, 2007.

[111] R. Rieger and J. Taylor. An adaptive sampling system for sensor nodes in body area networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17:183–189, 2009.

[112] E. Bashan, R. Raich and Alfred O. Hero III. Adaptive sampling: Efficient search schemes under resource constraints. *Communications and Signal Processing Lab, Univ. of Michigan, Ann Arbor, MI*, 2009.

[113] E. Candès and M. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25:21–30, 2008.

[114] H. Garudadri and P. Baheti. Packet loss mitigation for biomedical signals in healthcare telemetry. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 2450–2453, 2009.

[115] R. Thomas and et al. An electrocardiogram-based technique to assess cardiopulmonary coupling during sleep. *Sleep-New York Then Westchester*, 28:1151, 2005.

[116] J. Mietus. HRV in sleep apnea detection and sleep stability assessment at HRV [Online]. Available at: http://www.physionet.org/events/hrv-2006/mietus-2.pdf, [Accessed: Jan. 2015].

[117] E. Mazomenos, T. Chen, A. Acharyya, A. Bhattacharya, J. Rosengarten and K. Maharatna. A time-domain morphology and gradient based algorithm for ECG feature extraction. In *2012 IEEE International Conference on Industrial Technology*, pages 117–122, 2012.

[118] L. Sörnmo. Time-varying digital filtering of ECG baseline wander. *Medical and Biological Engineering and Computing*, 31:503–508, 1993.

[119] M. Raifel and S. Ron. Estimation of slowly changing components of physiological signals. *IEEE Transactions on Biomedical Engineering*, 44:215–220, 1997.

[120] R. Von Borries, J. Pierluissi and H. Nazeran. Wavelet transform-based ECG baseline drift removal for body surface potential mapping. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the. IEEE*, pages 3891–3894, 2005.

[121] P. Chazal, M. O'Dwyer and R. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51:1196–1206, 2004.

[122] S. Pei and C. Tseng. Elimination of AC interference in electrocardiogram using IIR notch filter with transient suppression. *IEEE Transactions on Biomedical Engineering*, 42:1128–1132, 1995.

[123] Y. Ider, M. Saki and H. Gucer. Removal of power line interference in signal-averaged electrocardiography systems. *IEEE Transactions on Biomedical Engineering*, 42:731–735, 1995.

[124] D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

[125] J. Talmon, J. Kors and J. Van Bemmel. Adaptive Gaussian filtering in routine ECG/VCG analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34:527–534, 1986.

[126] N. Thakor and Y. Zhu. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Transaction on Biomedical Engineering*, 18:785–794, 1991.

[127] P. Raphisak, S. Schuckers and A. Curry. An algorithm for EMG noise detection in large ECG data. In *Computer in Cardiology*, pages 369–372, 2004.

[128] B. Köhler, C. Hennig and R. Orglmeister. The principles of software QRS detection. *Engineering in Medicine and Biology Magazine, IEEE*, 21:42–57, 2002.

[129] J. Pan and W. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32:230–236, 1985.

[130] M. Elgendi, B. Eskofier, S. Dokos and D. Abbott. Revisiting QRS detection methodologies for portable, wearable, battery-operated, and wireless ECG systems [Online]. Available at: http://www.rxiv.org/pdf/1301.0058v2.pdf, [Accessed: Dec. 2013].

[131] P. Laguna, R. Jané and P. Caminal. Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database. *Computers and Biomedical Research (Now: Journal of Biomedical Informatics)*, 27:45–60, 1994.

[132] P. Chazal and R. Reilly. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 53:2535–2543, 2006.

[133] C. Panagiotou, S. Dima, E. Mazomenos, J. Rosengarten, K. Maharatna, J. Gialelis and J. Morgan. Detection of myocardial scar from the VCG using a supervised learning approach. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 7326–7329, 2013.

[134] J. Willems, P. Arnaud, J. Van Bemmel, P. Bourdillon, C. Brohet, S. Dalla Volta, J. Andersen, R. Degani, B. Denis and M. Demeester. Assessment of the performance of electrocardiographic computer programs with the use of a reference data base. *Circulation*, 71:523–534, 1985.

[135] B. Celler, P. Chazal and N. Lovell. Power spectral density estimates of populations of normal and abnormal 12 lead and Frank lead ECGs. In *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, pages 1407–1408, 1996.

[136] P. Chazal and B.Celler. Selection of optimal parameters for ECG diagnostic classification. In *Computers in Cardiology, IEEE*, pages 13–16, 1997.

[137] C. Li and C. Zheng. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42:21–28, 1995.

[138] J. Martínez, R. Almeida, S. Olmos, A. Rocha and P. Laguna. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51:570–581, 2004.

[139] P. Chazal and R. Reilly. A comparison of the use of different wavelet coefficients for the classification of the electrocardiogram. In *Proceedings 15th International Conference on Pattern Recognition*, pages 255–258, 2000.

[140] E. Jayachandran, P. Joseph and R. Acharya. Analysis of myocardial infarction using discrete wavelet transform. *Journal of Medical Systems*, 34:985–992, 2010.

[141] M. Llamedo and J. Martínez. Analysis of a semiautomatic algorithm for ECG heartbeat classification. In *Computing in Cardiology, IEEE*, pages 137–140, 2011.

[142] S. Rogers and M. Girolami. *A First Course in Machine Learning*. CRC Press, London, 2012.

[143] E. Alpaydin. *Introduction to Machine Learning*. The MIT Press, London, 2 edition, 2010.

[144] J. Hu. History of machine learning [Online]. Available at: http://www.aboutdm.com/2013/04/history-of-machine-learning.html, [Accessed: Jan. 2014].

[145] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958.

[146] J. Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.

[147] R. Lippmann. An introduction to computing with neural nets. *ACM SIGARCH Computer Architecture News*, 16:7–25, 1987.

[148] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 297:273–297, 1995.

[149] J. Zhang and et al. Modified logistic regression: An appSroximation to svm and its applications in large-scale text categorization. *International Conference on Machine Learning*, 2003.

[150] R. Fan, and et al. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[151] R. Polikar. Pattern recognition. *WILEY ENCYCLOPEDIA OF BIOMEDICAL ENGINEERING*, pages 1–22, 2006.

[152] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313:504–507, 2006.

[153] R. Neal. Bayesian methods for machine learning. *NIPS tutorial*, 13, 2004.

[154] O. Sayadi, and M. Shamsollahi. A model-based Bayesian framework for ECG beat segmentation. *Physiological Measurement* , 30:335, 2009.

[155] M. Wiggins, and et al. Evolving a Bayesian classifier for ECG-based age classification in medical applications. *Applied soft computing 8.1*, pages 599–608, 2008.

[156] S. Marsland. *Machine Learning: An Algorithmic Perspective*. CRC Press, London, 2009.

[157] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, pages 491–502, 2005.

[158] E. Frank. Weka 3: data mining software in Java (Lastest Stable Version 3-6-10) [Online]. Available at: http://www.cs.waikato.ac.nz/ml/weka/, [Accessed: Jan. 2014].

[159] E. Pekalska and R. Duin. PRTools: A Matlab toolbox for pattern recognition (Lastest Ver. PRTool5) [Online]. Available at: http://prtools.org/, [Accessed: Jan. 2014].

[160] C. Chang and C. Lin. LIBSVM – A Library for Support Vector Machines (Lastest Ver. 3.17) [Online]. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/, [Accessed: Jan. 2014.

[161] J. Han, M. Kamberi and J. Pei. *Data Mining : Concepts and Techniques*. Morgan Kaufmann Publishers, London, 3 edition, 2011.

[162] G. Magoulas and A. Prentza. Machine learning in medical applications. *Machine Learning and Its Applications*, pages 300–307, 2001.

[163] I. Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23:89–109, 2001.

[164] H. Banaee, M. Ahmed and A. Loutfi. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, 13:17472–17500, 2013.

[165] L. Zadeh. Fuzzy logic, neural networks, and soft computing. *Communications of the ACM*, 37:77–84, 1994.

[166] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, London, 2006.

[167] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[168] I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh. *Feature Extraction: Foundations and Applications*. Springer, London, 2006.

[169] S. Saitta. Standardization vs. Normalization [Online]. Available at: http://www.dataminingblog.com/standardization-vs-normalization/, [Accessed: Jan. 2014].

[170] H. Liu and H. Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, Boston, 1998.

[171] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand and H. Liu. Advancing feature selection research ASU Feature Selection Repository. Technical report, Arizona State University, 2014.

[172] G. John, R. Kohavi and K. Pfleger. Irrelevant features and the subset selection problem. In *Machine Learning: Proceedings of Eleventh International Conference*, pages 121–129, 1994.

[173] M. Dash and H. Liu. Feature Selection for Classification. *Intelligent Data Analysis*, 1:131–156, 1997.

[174] Y. Saeys, I. Inza and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23:2507–2517, 2007.

[175] D. Swets and J. Weng. Efficient content-based image retrieval using automatic feature selection. In *Computer Vision, 1995. Proceedings., International Symposium on*, pages 85–90, 1995.

[176] Y. Rui, T. Huang and S. Chang. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation*, 10:39–62, 1999.

[177] Y. Yang and J. Pedersen. A comparative study on feature selection in text categorization. In *International Conference on Machine Learning*, pages 412–420, 1997.

[178] K. Nigam, A. McCallum, S. Thrun and T. Mitchell. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39:103–134, 2000.

[179] E. Leopold and J. Kindermann. Text categorization with support vector machines. How to represent texts in input space? *Machine Learning*, 46:423–444, 2002.

[180] K. Ng and H. Liu. Customer retention via data mining. *Artificial Intelligence Review*, 3:569–590, 2000.

[181] G. Forman and C. Ira. Learning from little: Comparison of classifiers given little training. *Knowledge Discovery in Databases: PKDD*, pages 161–172, 2004.

[182] S. Sohn, W. Kim, D. Comeau, W. Wilbur. Optimal training sets for Bayesian prediction of MeSH® assignment. *Journal of the American Medical Informatics Association*, 15:546–553, 2008.

[183] S. Osowski and T. Linh. ECG beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering*, 48:1265–1271, 2001.

[184] T. Linh, S. Osowski and M. Stodoloski. On-line heart beat recognition using Hermite polynomials and neuron-fuzzy network. *IEEE Transactions on Instrumentation and Measurement*, 52:1224–1231, 2003.

[185] S. Osowski, L. Hoai and T. Markiewicz. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering*, 51:582–589, 2004.

[186] S. Mitra, M. Mitra and B. Chaudhuri. A rough set-based inference engine for ECG classification. *IEEE Transactions on Instrumentation and Measurement*, 55:2198–2206, 2006.

[187] F. Minhas and M. Arif. Robust electrocardiogram (ECG) beat classification using discrete wavelet transform. *Physiological Measurement*, 29:555–570, 2008.

[188] P. Chazal and R. Reilly. A comparison of the ECG classification performance of different feature sets. In *Computers in Cardiology*, volume 27, pages 327–330, 2000.

[189] L. Bao. Optimal classification [Online]. Available at: https://onlinecourses.science.psu.edu/stat557/node/38, [Accessed: Jan. 2014].

[190] A. Prugel-Bennett. Optimisation: gradient descent, quadratic minima, differing length scales [Online]. Available at: https://secure.ecs.soton.ac.uk/notes/comp3008/lectures/apb-lectures/, [Accessed: Jan. 2014].

[191] W. Noble. What is a support vector machine? *Nature Biotechnology*, 24:1565–1567, 2006.

[192] C. Hsu, C. Chang and C. Lin. A practical guide to support vector classification. Technical report, National Taiwan University, 2003.

[193] D. Hand. Assessing the performance of classification methods. *International Statistical Review*, 80:400–414, 2012.

[194] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, pages 1137–1145, 1995.

[195] W. Jamal and et al. Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates. *Journal of Neural Engineering*, pages 1–27, 2014.

[196] E. Mazomenos, D. Biswas, A. Acharyya, T. Chen, K. Maharatna, J. Rosengarten, J. Morgan and N. Curzen. A low-complexity ECG feature extraction algorithm for mobile healthcare applications. *IEEE Journal of Biomedical and Health Informatics*, 17:459–469, 2013.

[197] R. Almeida, J. Martínez, A. Rocha and P. Laguna. Multilead ECG delineation using spatially projected leads from wavelet transform loops. *IEEE Transactions on Biomedical Engineering*, 56:1996–2005, 2009.

[198] A. Ghaffari, M. Homaeinezhad, M. Akraminia, M. Atarod and M. Daevaeiha. A robust wavelet-based multi-lead Electrocardiogram delineation algorithm. *Medical Engineering & Physics*, 31:1219–1227, 2009.

[199] N. Thakor, J. Webster and W. Tompkins. Estimation of QRS complex power spectra for design of a QRS filter. *IEEE Transactions on Biomedical Engineering*, 11:702–706, 1984.

[200] P. Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38:910–917, 1992.

[201] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley. PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101:e215–e220, 2000.

[202] P. Laguna, R.G. Mark, A. Goldberg and G.B. Moody. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computer in Cardiology*, volume 24, pages 673–676, 1997.

[203] Southampton General Hospital. Blood, heart and circulation [Online]. Available at: http://www.uhs.nhs.uk/OurServices/Bloodandcirculation/Bloodandcirculation.aspx, [Accessed: Jan. 2014].

[204] CSE Working Party. Recomendations for measurement standards in quantitative electrocardiography. *European Heart Journal*, 6:815–825, 1985.

[205] F. Rincón, J. Recas, N. Khaled and D. Atienza. Development and evaluation of multilead wavelet-based ECG delineation algorithms for embedded wireless sensor nodes. *IEEE Transactions on Information Technology in Biomedicine*, 15:854–63, 2011.

[206] R. Jane, A. Blasi, J. Garcia and P. Laguna. Evaluation of an automatic threshold based detector of waveform limits in Holter ECG with the QT database. *Computers in Cardiology*, 24:295–299, 1997.

[207] V. Madisetti. *The Digital Signal Processing Handbook*. CRC Press, London, 2 edition, 2010.

[208] O. Rosso a, S. Blanco, J. ordanova, V. Kolev, A. Figliola, M. Schürmann and E. Baçar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105:65–75, 2001.

[209] G. McDarby, B. Celler and N. Lovell. Characterising the discrete wavelet transform of an ECG signal with simple parameters for use in automated diagnosis. In *Bioelectromagnetism, 1998. Proceedings of the 2nd International Conference on*, pages 31–32, 1998.

[210] A. Khazaee and A. Ebrahimzadeh. Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features. *Biomedical Signal Processing and Control*, 5:252–263, 2010.

[211] F. Mörchen. Time series feature extraction for data mining using DWT and DFT. Technical report, Philipps-University Marburg, 2003.

[212] S. Mallat. *A Wavelet Tour of Signal Porcessing: The Sparse Way*. Academic Press, London, 3 edition, 2009.

[213] F. Badilini, T. Erdem, W. Zareba and A. Moss. ECGScan: a method for conversion of paper electrocardiographic printouts to digital electrocardiographic files. *Journal of Electrocardiology*, 38:310–318, 2005.

[214] Y. Sun and K. Weber. Infarct scar: a dynamic tissue. *Cardiovascular Research*, 46:250–256, 2000.

[215] B. Drew, M. Pelter, D. Brodnick, A. Yadav, D. Dempel and M. Adams. Comparison of a new reduced lead set ECG with the standard ECG for diagnosing cardiac arrhythmias and myocardial ischemia. *Journal of Electrocardiology*, 35 Suppl:13–21, 2002.

[216] M. Hossain, T. Aziz and M. Haque. ECG compression using multilevel thresholding of wavelet coefficients. In *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 321–326, 2008.

[217] D. Biswas, E. Mazomenos and K. Maharatna. ECG compression for remote healthcare systems using selective thresholding based on energy compaction. In *2012 International Symposium on Signals, Systems, and Electronics (ISSSE)*, pages 1–6, 2012.

[218] M. Grčar and et al. kNN versus SVM in the collaborative filtering framework. *Data Science and Classification*, pages 251–260, 2006.

[219] J. Bland. *An Introduction to Medical Statistics*. Oxford University Press, London, 3 edition, 2000.

[220] J. Bland and D. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327:307–310, 1986.

[221] R. Fensli, T. Gundersen, T. Snaprud and O. Hejlesen. Clinical evaluation of a wireless ECG sensor system for arrhythmia diagnostic purposes. *Medical Engineering & Physics*, 35:697–703, 2013.

[222] J. Bland and D. Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8:135–160, 1999.

[223] G. Myatt and W. Johnson. *Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications*. WILEY, New Jersey, 2009.

[224] B. Hanzon. The area enclosed by the (Oriented) Nyquist diagram and the Hilbert-Schmidt-Hankel Norm of a linear system. *IEEE Transactions on Automatic Control*, 37:835–839, 1992.

[225] A. Acharyya, K. Maharatna, B. Al-Hashimi and J. Reeve. Coordinate rotation based low complexity N-D FastICA algorithm and architecture. *IEEE Transactions on Signal Processing*, 59:3997–4011, 2011.

[226] I. Romero and L. Serrano. ECG frequency domain features extraction: A new characteristic for arrhythmias classification. In *Proceedings of the 23rd Annual EMBS International Conference*, pages 2006–2008, 2001.

[227] N. Acır. Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm. *Neural Computing and Applications*, 14:299–309, 2005.

[228] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive Computing*, pages 1–17, 2004.

[229] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[230] DMML Group. Feature selection algorithms at Arizona State University [Online]. Available at: http://featureselection.asu.edu/software.php, [Accessed: Jan. 2014].

[231] K. Kira and L. Rendell. The feature selection problem: Traditional methods and a new algorithm. In *Association for the Advancement of Artificial Intelligence*, pages 129–134, 1992.

[232] I. Kononenko. Estimating attributes: analysis and extensions of RELIEF. In *Machine Learning: ECML-94*, pages 171–182, 1994.

[233] T. Cover and J. Thomas. *Elements of Information Theory.* WILEY, London, 2 edition, 2006.

[234] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *International Conference on Machine Learning (ICML)*, pages 856–863, 2003.

[235] M. Hall and L. Smith. Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In *FLAIRS Conference*, pages 235–239, 1999.

[236] E. Rich and K. Kevin. *Artificial Intelligence.* McGraw-Hill, New York, 1991.

[237] H. Liu and H. Motoda. Feature transformation and subset selection. *IEEE Intelligent Systems and Their Applications*, 13:1–9, 1998.