IDENTIFYING THE USAGE ANOMALIES FOR ECG-BASED HEALTHCARE BODY SENSOR NETWORKS

Lei Chen

Doctor of Philosophy

University of York

Computer Science

November, 2016

Abstract

This thesis is looking into the dependability of a Electrocardiogram(ECG) based Healthcare Body Sensor Network system (HC-BSNs). For these type of devices, the dependability is not only depending on the devices themselves, but also heavily depending on how the devices are used. Existing literature has identified that there are around 4% of usage issues when existing ECG devices are used by professionals. The rate of usage issue will not be better for the ECG-Based HC-BSNs as these devices are more likely to be used by untrained people. Subsequently, it is with paramount importance to address the usage issues so that the overall dependability of the ECG-Based HC-BSNs can be assured. Our approach to address the usage issue is to detect the usage-related anomaly, which is contained in the captured signal when erroneous usage is made, and identify the cause to the usage-related anomaly automatically and without human intervention. By doing this, the user can be prompted with clearer and accurate correction instruction. Subsequently, the usage issues can be well corrected by the user. Based on the above concept, in this thesis, we have studied the anomalous signals which can be caused by the usage issues. Two methodologies, names as *AID* and *FFNAID*, have been proposed and evaluated to detect the usage-related anomalies. We have also studied how each usage issue can affect the signals on a mote, and we use the knowledge learnt from the study to propose a methodology, named as *ACLP*, to identify the root cause to the usage-related anomaly. All these methodologies are fully automated and does not require any human intervention once they are deployed. The evaluations have also shown the effectiveness of these methodologies.

Contents

viii

List of Figures

List of Tables

Acknowledgements

To the finish of this thesis, I would like to give my most sincere thanks to my supervisor, Dr. Iain Bate, for his great effort and time over the past four years. Dr. Iain Bate has not only provided a great technical guide for this research but also taught me how to carry out the academic research to the standard in the U.K. Except for the great effort he has put down to this research, he has also given me and my family some great help during our life in U.K. Without his tremendous support and effort, I could not enjoy my life during the past four years. I sincerely hope that this thesis can be accepted as an evidence of how great a supervisor he is.

I would also like to thank for my family including my parents and parents in law, my wife, and my little boy. Over the past four years, it is your great support to let me continue and finish my research and subsequently this thesis. The most special thanks I would like to give is to my wife, Siyu Wang. During these years, we fell in love, we got married, and we had our cutest little boy, Ferdinand. During my hardest time, it is always you who gave me endless support. All the things we have gone through will only make us become an even tighter family.

Another special thanks I would like to give to Dr. Xiaoyun Wang, who is the cardiologist that have been frequently consulted to in this research. She has no obligation to participate this research. However, no matter what question I have related to the electrocardiogram, she had always been there to help me without any complaints. Finally, I would like to thank all the people that have been in my life during the past four years. Due to the page limitation, I will not name everyone here. However, thank you very much for appearing in my life and bringing the happiness, sadness, furiousness, and calmness. All of these have composed to my wonderful life.

Declaration

The work in this thesis has been carried out by the author between October 2012 and September 2016 at the Department of Computer Science, University of York. This work has not previously been presented for an award at this, or any other, University. Apart from work whose authors are clearly acknowledged and referenced, all other works contained in this thesis represents the original contribution of the author. Some of the works and their results have been published by the author. The list of the publication is as follows:

- L. Chen, and I. Bate. Identifying usage anomalies for ECG-based sensor nodes. In Proceedings of the 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pages 77-82, IEEE, 2016.

Chapter 1

INTRODUCTION

It has been widely acknowledged that nowadays healthcare services are under great pressure. To stop the situation becoming worse or to improve the future healthcare services, various proposals have been proposed in the past decade. With the software, microprocessor, sensor and actuator technologies that have been evolving over the past twenty or so years, the proposal of using wireless connected embedded systems to provide healthcare services has attracted tremendous research interests [1]. In this thesis, a research addressing the usage issues of the proposed medical devices is presented. More specifically, the correct usage of those proposed medical devices is with paramount importance in order to make the captured signal being useful. However, it is hard to guarantee the correctness of the usage no matter for trained professionals or amateur caregivers. This thesis is going to present a method which can detect the anomalous signals caused by erroneous usages, and identify the erroneous usages. Such method is capable to run on a mote at run-time so that the user can be notified with the incorrect usage without delay while the overhead of communication and resource consumption (RAM, ROM and energy) is minimised. To begin with, this chapter will introduce the concept of the proposed future medical devices and the current research in this area. By doing these, the research presented in this thesis will be motivated.
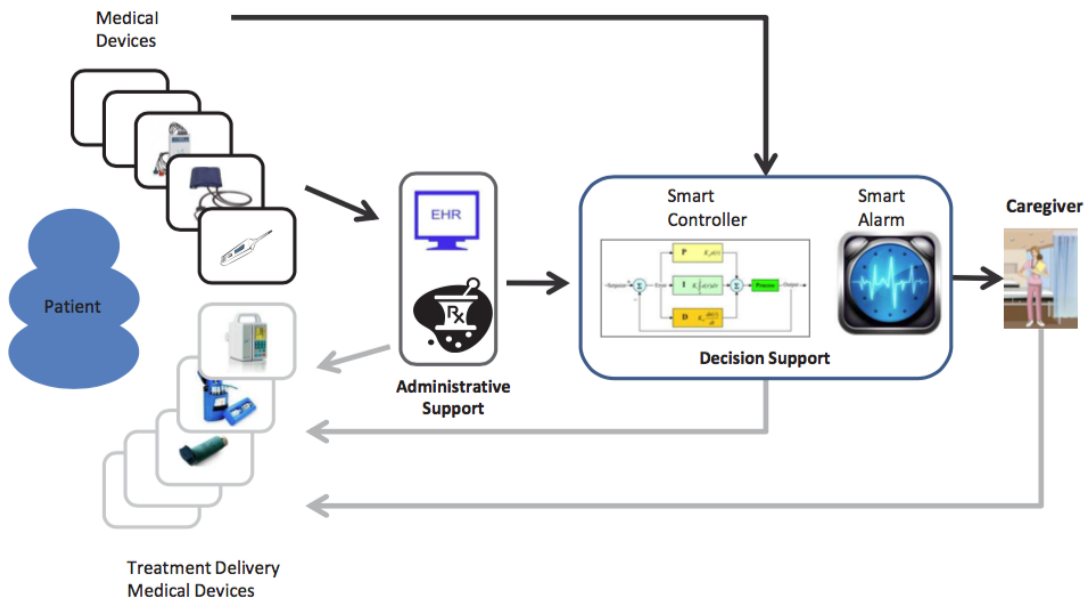
Figure 1.1: The concept of Medical Cyber-Physical Systems (quoted from [45])

1.1 Concept

1.1.1 Medical Cyber-Physical System

Cyber-Physical System (CPS) is described as "physical and engineered systems whose operations are monitored, controlled, coordinated and integrated by a computing and communication core" [66]. Such system forms a close-loop system that senses the physical world through sensors, makes decision through the processing devices and executes the decision through actuators [79]. It has been discussed that the CPS will change how the humans interact with the physical world in the future, just like how the internet has already changed the way that people interacts with each other [66]. However, the description of CPS sounds familiar as the term "embedded system", which has been used for decades, and can represent exactly the same functionalities [43]. It has been summarised that the key step from embedded system to CPS is to network those embedded devices [43] especially using wireless technology [64].

It has been proposed that by adapting the concept of CPS into the development of future medical devices, it is possible to maintain or improve the future healthcare service quality even with the pressures from high healthcare cost, ageing

populations, and diminishing medical professional resources [44]. Such systems are called Medical Cyber-Physical System (MCPS). The medical devices can be categorised into physiological signal sensing devices (e.g Electrocardiogram), and medical treatment delivery devices (e.g infusion pump). Currently, most of the medical devices work in a standalone manner. Individually, they are open-loop systems. When they are deployed to clinical scenarios, the loop is closed by the caregiver who act as a controller. Although some of the recent medical devices can work collaboratively with other devices via wired connection, these medical devices are still designed to rely on human intervention [44]. With the MCPS, the system becomes a closed-loop autonomous system in which the caregivers are replaced by additional computational entities [45].

Figure 1.1 shows the concept of MCPS. In the figure, the Medical Devices are referred to the physiological signal sensing devices, which are deployed to patient. The captured data is then fed to Administrative Support and Decision Support Devices. The treatment decision from those support devices is then be executed by the Medical Treatment Delivery Devices. At the same time, the system has the ability to show the necessary information (e.g captured signals, decisions etc.) to the caregiver, who can either monitor the system as a risk-control measure [44] or intervene the operation of Medical Treatment Delivery Devices. By adapting MCPS in future healthcare service, the following benefits over existing medical devices have been concluded so far:

Easier deployment

For those current medical devices that have the ability to collaborate with each other, the deployment requires plugging in the correct cable to the correct interface. The storage and deployment of these devices becomes expensive in terms of time and cost [64]. By adapting MCPS in future healthcare service, it has been promised that the deployment of medical devices will be significantly easier than existing medical devices [64].

Better connectivity

As the current medical devices are highly proprietary [44], different devices from different vendors uses different cables, interfaces and protocols. Subsequently, The connectivity between medical devices from different vendors is suffered. The wired communication between devices will also result in a so-called "malignant spaghetti" situation with the number of connected devices increase [64], which simply means that the cables are messed up in an inseparable way. By getting rid of cables, the connectivity between devices can be improved. Furthermore, it has been proposed that if all medical devices are adapting the same communication protocols, which is called the Medical Device Plug-and-Play (MD PnP) for MCPS [26], the connectivity between devices from different vendors can be further improved.

Higher patient safety

Easier deployment and better connectivity will both contribute to the higher patient safety. Furthermore, as MCPS introduces additional computational entities to replace the caregivers, the human error can be eliminated if the software is well designed [7].

Better flexibility

Current medical devices require both patients and medical personnel being present in the treatment room [64]. By adapting MCPS, either the patient or medical personnel can have the flexibility to move around or to provide healthcare without compromising the service quality.

1.1.2   Healthcare Body Sensor Networks

It has been proposed that Wireless Sensor Networks (WSNs) technology has great potential to provide the ability of sensing the physical world for CPS in real-time [79]. Among those application domains of WSNs, a specific application area of WSNs is to deploy motes to human body. Such application area is sub-categorised into Body Sensor Networks (BSNs), which promise novel uses in healthcare, fitness and entertainment purposes [32]. As the continuously growing pressure of

existing healthcare service and the proposal of MCPS, a vast amount of researches have been concentrating on the healthcare application [32]. In this thesis, the application of BSNs for healthcare purpose is referred to Healthcare Body Sensor Networks (HC-BSNs). Figure 1.2 illustrates a concept of the HC-BSNs. In general, several motes are deployed around human body with a Body Area Data Aggregator receiving and relaying data to the remote devices or servers. Meanwhile, the data aggregator may have the ability to show the collected data or signal to either the monitoring target or caregivers.

For the motes used in HC-BSNs, they may have some differences compared to the motes used in a WSN system. In WSNs, motes are designed to perform the same functionality so that the deployment of the mote does not need to be engineered ahead of deployment. As the motes are deployed to a natural environment, it has less requirement for each mote in terms of the size and material [42]. On the other hand, the functionality of each mote in HC-BSNs application is different to others. Each mote have a specific signal to be sensed. For example, in Figure 1.2, six motes are deployed to the human body. Each of them senses a specific physiological signal (e.g ECG, Blood Pressure etc.). As each mote has a specific functionality, the location of each mote need to be carefully decided. Furthermore, as the motes are deployed onto human body, the size of the mote must be small enough so that the mote is highly wearable, and the mote must be bio-compatible [24].

So far, many HC-BSN prototypes have been proposed by other researchers. For example, Shnayder et al. designed a comprehensive platform for medical sensor networks called CodeBlue [75]. In this prototype, a range of medical sensors including ECG and pulse oximetry (SpO_2) were built based on existing motes. Besides the hardware, a software interface which is able to show the captured medical signal in real-time had also been implemented. Furthermore, by deploying several fixed location nodes, the software interface is able to display the possible location of each monitored patient.

Although CodeBlue is able to monitor the medical signals of a patient, there isn't
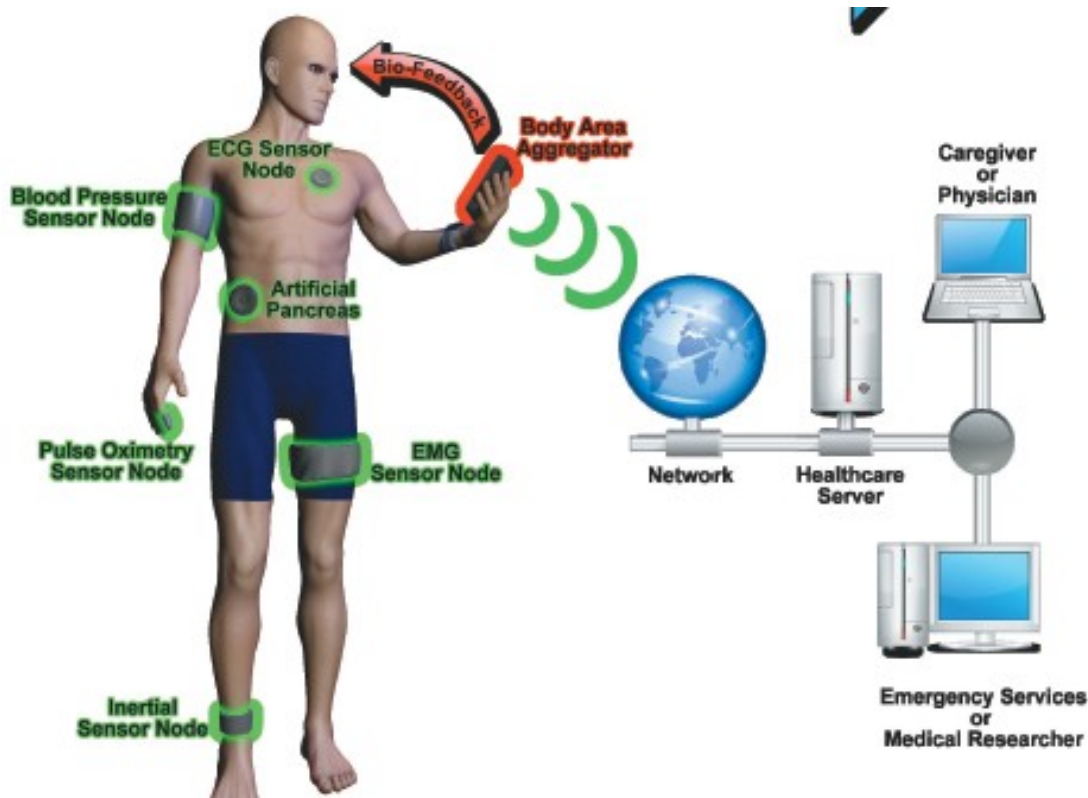
Figure 1.2: The concept of Healthcare Body Sensor Networks (quoted from [32])

any mechanism that can detect the abnormal condition and trigger the alert for caregivers. In contrast, AMON [6] is a similar system to CodeBlue but with a medical algorithm which can trigger the alarm when the collected data is in risk zone. The medical algorithm is a rule-based algorithm. The collected data is pre-categorised into three groups: normal, risk and high-risk. For each group, a range is set according to physiological and medical requirements. After each signal is collected, the algorithm compares the measurement to those groups' value range. If the collected data is in the risk or high-risk group, the system will collect other correlated measurements in order to confirm the risk. An overall index of risk is calculated based on the weighted measurements. If the index is still in the risk groups, the alarm will be triggered.

1.1.3 Healthcare Model

The proposal of MCPS and HC-BSNs has enabled the remote healthcare monitoring or treatment delivery. Subsequently, it has been proposed that the future healthcare model will be shifted from nowadays' centralised healthcare model to

distributed healthcare model [88].

Centralised Healthcare Model

Nowadays' healthcare model is described as the centralised, expert-driven, crisis-care model [88]. It means that the healthcare services are provided at a specific location (e.g hospital) by the trained professionals (e.g doctors or nurses). The main limitations for existing healthcare model, according to Aziz et al. [9], is that it requires the patient to frequently visit healthcare when it is necessary. However, the patient has less or no ability to perform self-diagnosis to decide whether it is necessary to go to healthcare service centre. Or, the patient relies on long term monitoring in the healthcare centre but not receiving the treatment during the time when the patient is in the healthcare centre. In these but not limited to these circumstances, a patient may visits or stays in the healthcare centre for unnecessary reasons (e.g the patient thinks he or she is not well). As a consequence, the healthcare service resources will be consumed rapidly, and the pressure on existing healthcare services becomes extremely heavy [1].

Distributed Healthcare Model

In contrast to centralised healthcare model, the distributed healthcare model does not fully rely on trained experts to provide the healthcare services. Instead, some common long-term monitoring can be achieved without the patient being present at the healthcare service centre. For example, the aforementioned AMON [6] is a remote ECG-Based healthcare monitoring system prototype. It can monitor a patient's vital signs and provide brief diagnosis to decide the risk level of the monitoring subject. By using this prototype, the patient will need significantly less visits to healthcare centre when the patient is in low risk states. At the same time, the risk state can be constantly monitored so the patient will not be in risk when the patient is not at the healthcare centre. By doing this, the pressure at the healthcare centre can be decreased. In the distributed healthcare model, the caregivers is not necessarily to be the trained professionals. It can be any informal caregivers (e.g family members, friends, neighbours etc.), who only need to aid the patient to use

or maintain the system instead of giving suggestions or performing diagnosis.

1.2   Dependability and Usage

For a system being used for healthcare purpose, the system itself must be dependable at the first place [79]. A system's dependability is defined by Avizienis et al. as "The ability to avoid service failures that are more frequent and more severe than is acceptable to the user" [8]. It is an encapsulation of the following attributes [8] with an example being given for each attribute using a HC-BSNs for monitoring purpose:

- *Availability* - The readiness for correct services. For example, the availability can be referred to the system being able to sense and monitor the relevant signals.

- *Reliability* - The continuity of correct services. For example, the reliability can be referred to how long the relevant signals can be sensed and monitored correctly before the system lost the ability.

- *Safety* - The absence of catastrophic consequences on the users and the environment. For example, a system can no longer monitor a patient's health status correctly. In this case, the safety can be referred to whether the patient will be leave in danger.

- *Integrity* - The absence of improper system alteration. For example, the integrity can be referred to whether the consumable materials used by the system (e.g ECG electrode pads) are all complying with the standard set by the system.

- *Maintainability* - The ability for process to undergo modifications and repairs. For example, when the system can no longer monitor the patient's status, the maintainability can be referred to whether the system can be repaired and how hard the repair will be.

For the proposed HC-BSNs to be dependable, the usage of the system can play an important role as various literatures [13, 72, 33] have discussed the consequence when a medical device is incorrectly used. To analyse how the system's usage can have impact to the system's dependability, an application must be chosen as "the dependability must be analysed with a focus on a particular service or application" [22]. In this thesis, a HC-BSNs with ElectroCardioGram (ECG) sensor is chosen as the example system. In the rest of this thesis, such system is named as ECG-Based HC-BSNs. The reasons why an ECG-Based HC-BSNs is chosen is due to the following aspects. Firstly, ECG is one of the widely used example systems not only in HC-BSNs research domain, but also in other research domains. As a result, many existing works can be referenced and existing resources that can be used in this research. Secondly, the use of ECG devices requires certain level of expertise, which is why ECG devices are currently used by trained professionals and huge efforts have been spent on the training. Also, the long term monitoring of ECG signal is as important as short term diagnosis as the emergencies like heart attack are best detected from ECG signals. Combining these two points makes the ECG sensors become the most representative case for this research. Such case is defined as the long-term physiological signal monitoring system that requires correct usage, but the correct usage is hard to guarantee due to the required expertise. Thirdly, the availability of relevant devices are important for this research. When this research is carried out, HC-BSNs with ECG sensor (Shimmer2r [2] with ECG daughter board) is the only available device.

For an ECG-Based HC-BSNs, it typically features many sensor pads that are each connected to the mote via a wire. For the signal to be of sufficient quality and useful, it is important that the sensors are attached in the correct place with respect to the heart, that the sensors are securely connected with medical tape so that the signal noise is in an acceptable level (e.g. static noise is below a certain level), and that the leads are then connected to the right port of the mote. The correct use of ECG devices has already been proved to be difficult for trained professionals as various literatures [13, 72, 33] have highlighted the erroneous usages of electro-

cardiogram. It has also been reported that the frequency of erroneous ECG usage at the outpatient clinic is round 0.4% whereas the frequency at the intensive care unit is around 4%. Considering the ECG-Based HC-BSNs is usually being operated by untrained caregivers, the frequency of erroneous usage will not be better than those frequency achieved by trained professionals. Furthermore, the usage of the ECG devices may become erroneous during the operation of the system. For example, the ECG sensor pads may detach from monitoring target during the monitoring process due to rubbing between the device and other objects (e.g arms or cloths). In this case, although the device is being correctly used initially, usage issue starts to appear during the operation.

Through correlating the usage with the attributes of dependability, the following points can be summarised. Firstly, the correct use of the ECG device has been proven to be hard for trained professionals and is believed to be even harder for untrained caregivers to use ECG-Based HC-BSNs. In the case where the device is used erroneously, none of the captured physiological signals will be useful to professionals. Consequently, the system can be considered as not being available to provide adequate services. Subsequently, the availability is suffered. Secondly, the usage of the ECG device may become erroneous after certain time of operation. If this is the case, the system can no longer provide the intended services after the point when the usage issue appears. That means the system's reliability will be suffered. Thirdly, it is even harder for the untrained caregivers to diagnose and rectify the erroneous usage they have made when using the ECG-Based HC-BSNs. Subsequently, it is also considered that the maintainability will be suffered in this case. As the availability, reliability, and maintainability can all be suffered in the case when an untrained caregiver operate the ECG-Based HC-BSNs, the dependability is therefore suffered as well. In order to assure an ECG-Based HC-BSN system's dependability, it is with paramount importance to eliminate the erroneous usages.

So far, to the best of the author's knowledge, none of the existing works in HC-BSNs area has looked into the usage issue. Nowadays, the operation of existing

ECG devices severely relies on the expertise of the operators. As a consequence, the main effort of addressing the usage issues is to improve the training quality of using those devices. However, there are some literatures which tried to rule out the usage issues in various medical environment through monitoring the usage and prompting the user with possible usage issues so that they can be rectified. For example, Simpson et al. [76] use a checklist approach to monitor the correctness of the usage of medical devices in Intensive Care Unit (ICU). Their experiment has shown that it can significantly improve the erroneous device usage rate. To tailor such approach for ECG-Based HC-BSNs, it is proposed that in order to eliminate the erroneous usage, the information of whether the device is being erroneously used and how to rectify the erroneous usage will need to be provided to the user so that the usage issue can be address even when the user is untrained caregivers. Based on the proposal, the first step is to understand how the erroneous usage can affect the captured signal and to detect usage-related anomaly which is defined as the signal that is corrupted by erroneous usage. This leads to the first research question:

> *Research Question 1: How the erroneous usage of the ECG-Based HC-BSNs can affect the captured signal and whether it is possible to detect the usage-related anomaly on a mote?*

After the usage-related anomaly is detected, the usage error need to be rectified so that the captured signal can be recovered. In the proposed distributed health-care model, as the informal caregivers are not necessarily to be the trained professionals , the level of expertise of operating ECG devices can vary dramatically. Consequently, it is barely possible that the informal caregiver can rectify the usage error only based on the information of the usage-related anomaly. In order to overcome this issue, it would be important to identify the root causes to the usage-related anomaly based on the signals on a mote environment and without human intervention. Subsequently, the information of the root cause and how to rectify it can be prompted to assist the informal caregivers to rectify the usage issues. This lead to the second research question:

Research Question 2: What is the relationship between the erroneous usages and the signals on a mote environment and whether the root cause to the usage-related anomaly can be identified using the signals on a mote environment and without human intervention?

The first and second research questions look into the possibility of detecting and identifying the usage anomalies on a mote so that no overhead in communication will be generated. However, as the proposed HC-BSNs are wirelessly connected, the off-mote resources (e.g extra processing power) may have positive impact to the system. To investigate the possible impact in terms of detection accuracy when adapting off-mote collaboration, the third research question in this thesis is as follow:

Research Question 3: Can the off-mote collaboration offer benefits in terms of the detection accuracy?

1.3   Thesis Hypothesis

Based on the motivation and research questions, the research work is carried out based on the following hypothesis:

Usage-related anomalies for ECG sensors can be detected and their causes identified in a reliable and resource efficient manner by comparing time domain signals based on recognised patterns from previous acceptable and anomalous signals.

1.4   Thesis Structure and Contributions

To present the research that has been done so far, the rest of this thesis contains five chapters. Among these five chapters, there are three technical chapters which individually address one of the three research questions. In this section, the structure of this thesis will be introduced and the contribution of each technical chapter will be summarised.

In Chapter 2, a literature review will be presented. Such review covers the knowledge of the electrocardiogram and the anomaly detection. Also, existing anomaly detection approaches will also be reviewed.

Chapter 3 will be the first technical chapter. It aims at address the *Research Question 1*. More specifically, the main contribution of this chapter is to demonstrate that the usage-related anomaly of an ECG-Based HC-BSNs can be accurately detected on a mote at run-time through comparing the captured signal to the pre-defined correct signal in time domain. Under this main contribution, the following contributions have also been made in this chapter:

- A methodology, named as *AID*, has been defined to demonstrate the ability to detect the usage-related anomaly on a mote at run-time

- The proposed methodology has been evaluated using ECG signals from different subjects to show its ability to detect the usage-related anomaly accurately regardless of the subjects

- The proposed methodology has been implemented on a mote and its overhead in terms of computation and memory has been evaluated to show the capability to run on a mote at run-time

With the usage-related anomaly having been detected, this research looks into the possibility of identify the root cause to the usage-related anomaly without human intervention in Chapter 4, which is the second research question for this research. On the way to address this research question, the following contributions have been made in this chapter:

- This research being the first research to study the effect that the erroneous usage can make to the on mote signals especially the signals from on-board accelerometer. Such study is necessary for identifying the root cause to the usage-related anomaly.

- Based on the study, a methodology inspired by the existing checklist approach adapted by cardiologist has been defined

- The proposed methodology has been evaluated using data recorded from voluntary participants and the result has demonstrated that the root cause to the usage-related anomaly can be identified without human intervention

In Chapter 5, the benefits of adapting off-mote collaboration, which is the third research question, have been demonstrated. Such work is inspired from the limitation of the previously proposed *AID* which is identified through the evaluation. The following contributions have been made during the research presented in this chapter:

- Proposing a new methodology called *FFNAID* to investigate the benefits of adapting off-mote collaboration

- Through evaluating the *FFNAID* using signals from different subjects, the effect of different configuration of the *FFNAID* has been demonstrated

- It has also demonstrated that if the *FFNAID* is properly configured, it can provide significant benefit in terms of detection accuracy compared to *AID*

- The evaluation has also demonstrated that sharing data between different subject has no obvious benefit in terms of detection accuracy

In the last chapter, this thesis will be concluded with the results from each technical chapters. This chapter will review the research questions and see how well they have been addressed. Finally, some possible future works in this area will be proposed.

Chapter 2

LITERATURE REVIEW

In order to investigate the usage issues of ECG-based HC-BSNs, this chapter will present a literature survey of the knowledge related to the electrocardiogram and anomaly detection. The main purposes is to provide a comprehensive knowledge to the problems this research is facing. The rest of this section is arranged as follows: Section 2.1 will present a survey of electrocardiogram. Such survey includes the type of ECG devices (Section 2.1.2), the deployment of the ECG recorder (Section 2.1.3), and the normal ECG signal composition (Section 2.1.4). The dependencies between the usage-related anomalies and their possible root causes will also be reviewed in Section 2.1.5. Then, Section 2.2 will present a literature survey of the anomaly detection. It includes the basic knowledge of the anomaly detection and existing anomaly detection approaches for ECG signals. Also, how existing works are evaluated will also be reviewed. At last, the finding from this chapter will be summarised.

2.1   Literature Review of Electrocardiogram

The history of electrocardiogram can be traced back to the beginning of 20th century. It is described that the introduction of chest X-Ray and electrocardiogram has enabled the human to explore the structure and function of the human heart [5]. Since then, the electrocardiogram has evolved continuously and has become not

only the most widely used but also cost-effective and non-invasive approach to diagnose the heart disease [23]. Such subject is known as cardiology nowadays. As this thesis focuses on the usage issues of ECG devices, it is necessary to understand the basic knowledge before the rest research can be carried out. Thus, this section is going to present a literature review of electrocardiogram.

2.1.1 ECG Signal Acquisition

The earliest electrocardiogram used a string galvanometer to record the potential electrical difference between the extremities [5]. Such electrical differences are the result of the heart muscle contraction which can also be known as the heart beats. With over a hundred years of evolution, the equipments to record the ECG signal have become smaller, more convenient to use, and more information from the heart can be measured. In general, modern ECG devices record the signals digitally. That means the electrical difference on the patient's skin is measured by an analogue-to-digital converter (ADC) at a fixed interval. Such interval between measurements is knowns as the sampling interval or sampling frequency. Existing work [65] has summarised that the sampling frequency of the ECG device can be as low as 125 Hz, at which frequency the signal's quality may be suffered under certain conditions. Therefore, it is less used in real applications. The recommended and widely used sampling frequency is 250 Hz and above as there is no significant difference can be observed between the captured signals whose sampling frequencies are higher than 250 Hz [65].

For the signals to be picked up by the ADC, it is necessary to amplify the analogue signals on the skin. The analogue amplification is usually referred to Analogue Front-End (AFE). According to the application report published by Texas Instruments [78], there are typically two approaches for ECG signal acquisition. The first approach is a high voltage gain AFE (e.g voltage gain of 500 for 2.5V ADC full-scale measurement and 5mV input signal) combined with a low resolution ADC (less than 16-bit). The second approach is to use a lower voltage gain AFE (e.g voltage gain of 5 for 2.5V ADC full-scale measurement and 5mV input signal) combined with a high resolution ADC (e.g 24-bit resolution). The performance of both approaches

in terms of the signal acquisition quality are similar. However, to achieve similar performance, the price to pay is different for each approach. For the first approach, attention is paid to minimise the noise whereas the second approach requires a significantly more expensive ADC compared to the first approach. In the context of ECG-based HC-BSNs, the first approach is usually used due to the constrained onboard environment. However, the AFE is usually not able to handle the noise well due to the low cost of the sensor.

2.1.2 Types of Electrocardiogram Devices

So far, many ECG measurement devices have been proposed for various purposes. Through summarising these devices, it can be categorised into four types. The main difference between different types is the number of leads. In this section, each type will be briefly introduced. At the end of this section, the device used in this research will also be introduced.

1-Lead ECG Recorder

First type of ECG devices is called 1-Lead ECG recorder. It is primarily used for basic heart monitoring, or simple educational or research purpose. However, there is a possibility of using 1-Lead ECG recorder to accomplish full 12-Lead ECG recordings in a sequential manner [27]. As there is only one lead, the deployment is very simple and the size of the device can be very pervasive. Subsequently, it is widely used for recording ECG signal during exercise [27]. Also, due to its simplicity and the ability for basic heart monitoring, many existing HC-BSNs have also adapted the 1-Lead ECG. For example, the ECG monitoring in the aforementioned CodeBlue [75] and AMON [6] are the 1-Lead ECG recorder.

3-Lead ECG Recorder

One of the widely used types of ECG Recorder is called 3-Lead ECG recorder. Although it is named as 3-Lead ECG, there is actually four electrodes to be placed on each of the limbs. As illustrated in Figure 2.1, the measurements are recorded between two electrodes. As there is one electrode being used for the reference
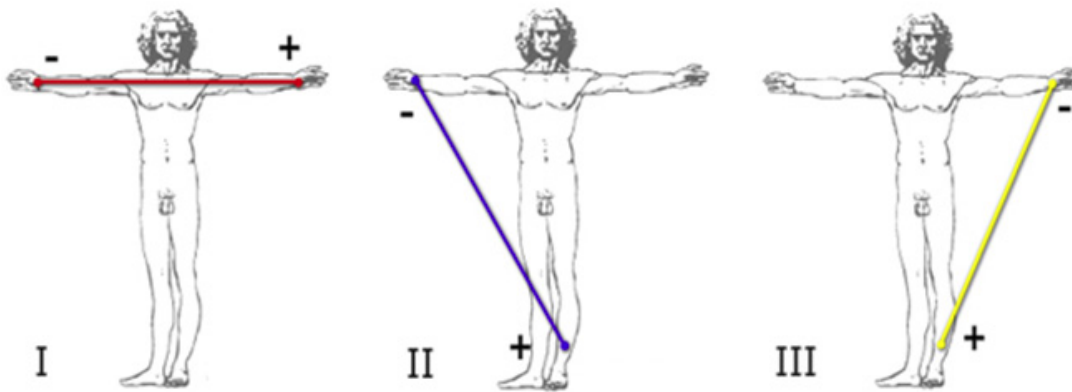
Figure 2.1: An illustration of 3-Lead ECG (Quoted from [3])

purpose (the electrode on right leg), any two of the rest three electrodes will provide one measurement. As a result, there are three measurements in total and they are named as Lead-I, Lead-II, and Lead-III. This type of ECG recorder is widely for continuous monitoring of heartbeat, heart rate and used for assessing arrhythmias [5]. However, there are still some 'silent areas' in the heart where the 3-Lead ECG can not measure [16].

12-Lead ECG Recorder

Due to those 'silent areas', which can not be measured using 1-Lead or 3-Lead ECG recorder, the modernest ECG recorder adapts the 12-Lead ECG recording approach. This type of ECG recorder is used in the clinical settings today. It provides a comprehensive view of the heart from different angles. The cardiologist can use the 12-Lead ECG recording to determine the exact location of the abnormality in a patient's heart.

Holter Recorder

A specific type of recorder is called Holter recorder which is used to provide the long term monitoring of the heart behaviour. As some of the heart conditions may happen at random time with random duration, it is not guaranteed that the normal ECG measurement will capture the abnormal signals. Subsequently, it is necessary to have a system that can monitoring the ECG signal in long-term. For the Holter recorder, as it is usually used for the long-term diagnosis, it usually adapts

the 12-Lead recording approach. However, there are also some 3-Lead Holter recorder.

For the Holter recorder, it usually measures the ECG signals and stores the measured signals to an on-board memory. When the prescribed recording duration is ended, the patient will hand the device back to professionals who will then off-load the data for diagnosis.

ECG-Based healthcare body sensor networks

The newest innovation of the traditional clinical ECG recorders is the ECG-Based healthcare body sensor networks (HC-BSNs). These type of ECG recorder can wirelessly transmit the recorded signal from the device to a designed destination. For example, the patient can wear the device at home and the captured signal can be transmitted to the doctor's computer for diagnosis.

In this research, the shimmer2r mote platform with the ECG daughter board [2] is used as the ECG-Based HC-BSNs. For this combination, it adapts the 3-Lead ECG recording approach. There is an on-board data storage (SD-Card) to temporary store the captured signal. There is also a radio communication device which can wirelessly communicate to other compatible devices. An accelerometer has also been integrated into the device so the motion status of the mote and the patient who wears the mote can be measured. The whole device is powered by an on-board battery.

2.1.3   The Deployment of ECG Recorder

In order to investigate the usage issue of the ECG-Based HC-BSNs, this section is going to present a review of how the ECG device should be used. As the device used in this research adapts the 3-Lead ECG recording configuration, this review will only cover the deployment of 3-Lead ECG recorder. Then, as each ECG device manufacturer may have some special requirement in terms of the deployment, the special deployment published by Shimmer in their ECG User Guide v1.11 [68] will also be introduced.
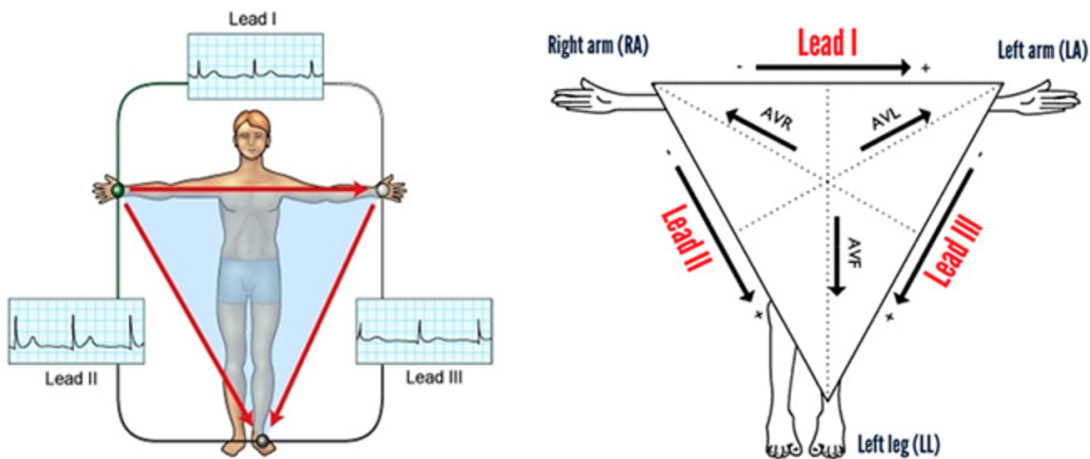
Figure 2.2: An illustration of the Einthoven's Triangle principal (Quoted from [3])

For the ECG recorder with 3-Lead ECG configuration, it features four bipolar limb electrodes, which are named as LA, RA, LL, and RL. Among these four limb electrodes, the RL is being used as the reference ground. For the rest three electrodes, they need to be placed based on the Einthoven's Triangle principal [59, 51, 28], which has been illustrated in Figure 2.2. Such triangle principal is based on the location of the heart and the angle in respect to the heart. The location of each electrode only need to follow the angle in respect to the heart and is not necessarily to be placed on the limbs. Subsequently, Shimmer recommends to place the four bipolar limb electrodes as shown in Figure 2.3. It can be seen from this figure that the three electrodes, RA, LA, and LL, form a Einthoven's Triangle and the RL is far away from the triangle as it is the reference ground. Based on this placement, the measurements of the three ECG Lead are as below:

- Lead I = LA-RA

  Lead I is the ECG vector signal from the RA to LA.

- Lead II = LL-RA

  Lead Ii is the ECG vector signal from the RA to LL.

- Lead III = LL-LA

  Lead Ii is the ECG vector signal from the LA to LL.
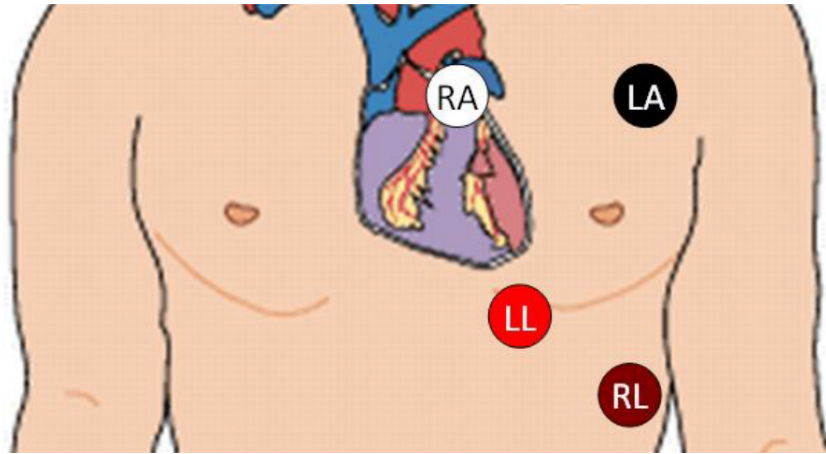
Figure 2.3: The placement of the four bipolar limb electrodes recommend by Shimmer (Quoted from [68])

On top of the placement of the electrodes, Shimmer has also made the following recommends in the ECG User Guide in order to assure the signal quality:

- Ensuring a good electrode to skin contact so that the likelihood of signal interference can be minimised

- Secure the cables, which connect between the electrodes and the mote, and the mote so that the noise from motion artefact can be minimised, and minimise the likelihood of skin contact issue due to the cable and mote pulling electrodes

With the correct deployment of the mote and electrodes, the ECG signal can be accurately recorded. Next section is going to explain the recorded correct ECG signal.

2.1.4 ECG Signal Composition

Over the years of research in the cardiology, the composition of the correct ECG waveform has been well defined. Figure 2.4 has illustrated an example signal with the annotation to the useful feature composition. As seen from Figure 2.4, an ECG waveform is generally composed of six key features, which are named as P-Wave, PR Segment, QRS Complex, ST Segment, T-Wave, and U-Wave. Among these

features, the U-Wave is not usually visible for every person [36]. Subsequently, it is usually ignored. For the rest of the features of an ECG waveform, Table 2.1 has shown their name, duration, and the heart activity reflected by the feature. Beside these feature composition, there is another key feature from the ECG signal, which is the Heart Rate (HR). The HR is usually derived by measuring the duration between two adjacent R-Peaks which is the highest peak in the QRS Complex. The interval between two adjacent R-Peaks is also known as RR-Interval. For the ECG signals with different heart rate, the length of RR-Intervals is changed but the duration of those ECG features remain unchanged unless the heart activity is changed (heart conditions) [59, 51, 28, 36]. As a result, there is an interval where the ventricles are relaxing and no heart activity is performed during this interval. This interval is referred to the Non-Cardio Activity Duration as illustrated in the Figure 2.4.

For an ECG signal to be useful to the cardiologist, the minimum requirement to the signal quality is that all the features that composed to the ECG waveform should be clear enough to the cardiologist's standard. Through discussing with a cardiologist, such standard may vary between different different cardiologists. In the next section, how the erroneous usage of the ECG device may make the signal become less useful to the cardiologist will be reviewed and their dependency will be defined using Fault Tree Analysis [46].

Feature	Duration	Heart Activity
P-Wave	80-100 ms	Atrial Depolarisation
PR Segment	120-200 ms	AV nodal delay
QRS Complex	80-120 ms	Ventricular Depolarisation
ST Segment	70-80 ms	Ventricles contracting and empty
T-Wave	Around 200 ms	Ventricular Repolarisation

Table 2.1: The detail feature composition of an ECG waveform with their name, duration, and which heart activity causes the feature

2.1.5 Dependencies Between the Usage Anomalies and Their Root Causes

Through the review of existing literature of how the usage can affect the captured signal from the ECG devices, three types of usage-related anomalies have been
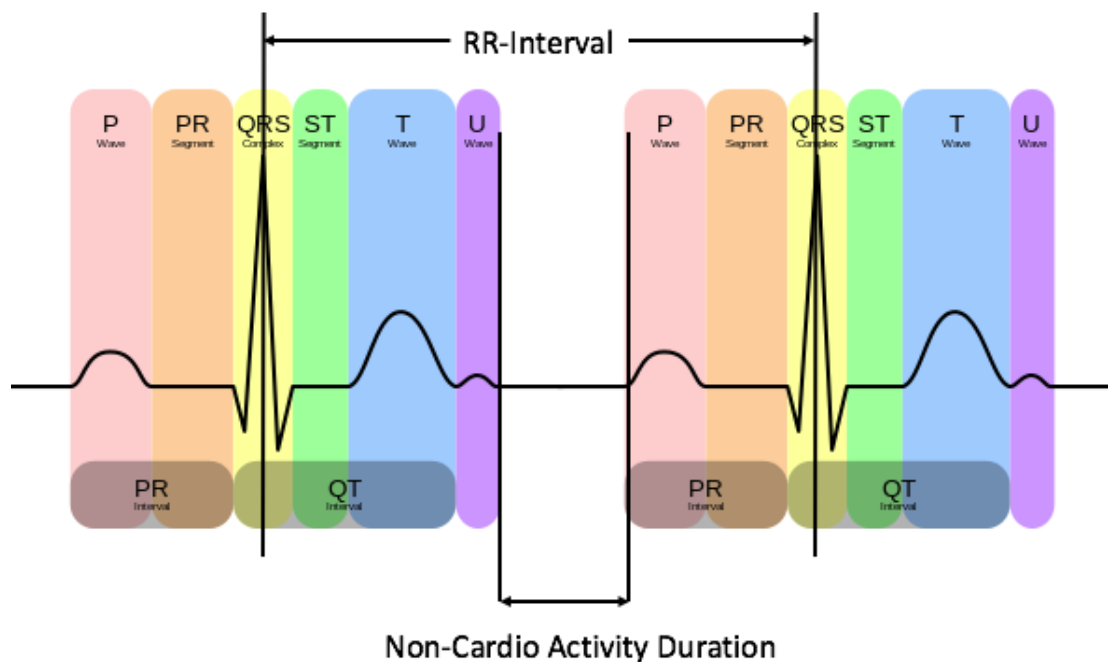
Figure 2.4: An example illustrates the composition of a correct ECG waveform

summarised: *Blank Signal* anomaly, *Inverted Signal* anomaly, and *Noisy Signal* Anomaly. At the same time, the root causes to the usage-related anomalies have also been reviewed. In order to clearly represent the dependency between the usage-related anomaly and the root cause to it, fault tree analysis [46] will be carried out to the finding from existing literatures. The overall derived fault tree is shown in Figure 2.5 and an illustration of each type usage-related anomaly has been shown in Figure 2.6. In the rest of this section, each usage-related anomaly will be discussed in detail.

Inverted Signal Anomaly

According to existing literatures [33, 71, 52], the cause to the *Inverted Signal* anomaly can only be the electrodes location being swapped. More specifically, as the ECG signal is a polarised signal, the swap of the electrodes location will lead to the signal being inverted. For example, Harrigan et al. [33] have stated that the most common electrodes location swap is the right and left arm electrodes being swapped. It will result in the signal on Lead-I being inverted. There is no other cause to the *Inverted Signal* anomaly can be found from existing literature.
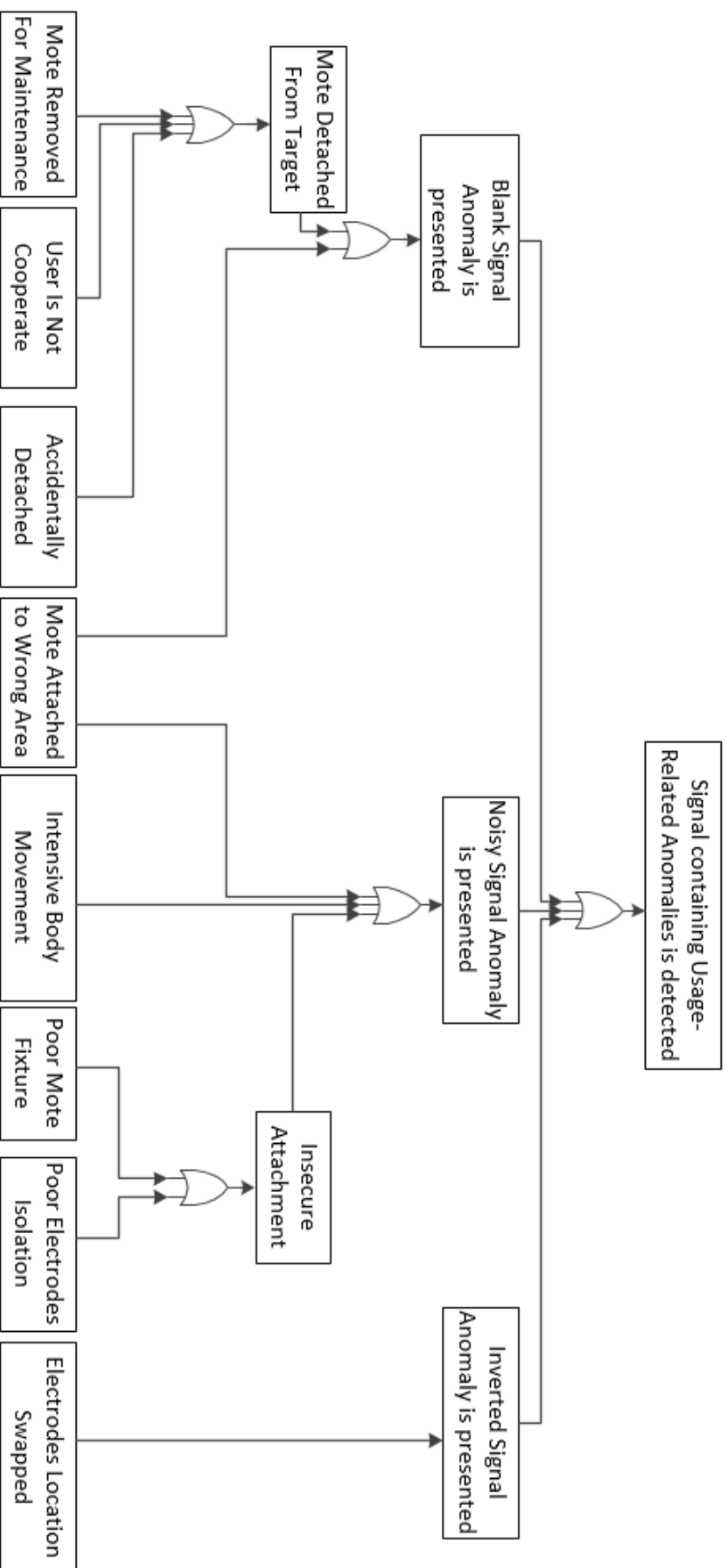
Figure 2.5: The derived fault tree for the usage-related anomalies and their root causes
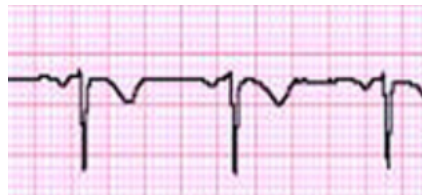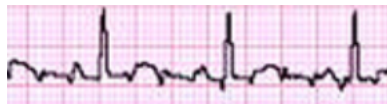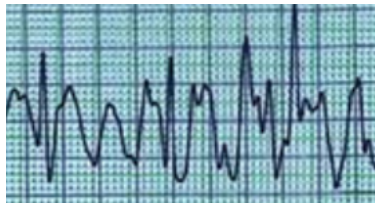
Figure 2.6: An illustration of the feature of each type of usage-related anomaly (all figures are quoted from [33]). (a) An Illustration of the feature of *Inverted Signal* anomaly (b) An Illustration of the feature of *Blank Signal* anomaly (c) An Illustration of the feature of still useful signal (d) An Illustration of the feature of *Noisy Signal* anomaly

For the *Inverted Signal* anomaly, the main feature, as the name implied, is the inverted P-QRS-T waveform [33]. As illustrated in Figure 2.6a, the captured signal is inverted compared to the correct signal illustrated in Figure 2.4.

Blank Signal Anomaly

Through the discussion with a cardiologist, the cause to the *Blank Signal* anomaly can be the mote being detached from the target or the mote being deployed to wrong area. Then, due to the way the ECG-based HC-BSNs is used, it is necessary to find out how the mote detaches from the target. In the fault tree, three sub-causes to the mote detachment have been concluded. First, the mote is removed for maintenance (e.g charging the battery). Second, the mote is accidentally detached from the target due to many possible reasons. The reason to the accidental detachment is out of the scope of this research so it will not be further broken down. The third cause is the mote is removed from the target for no obvious reason. In this research, such cause is referred to the user being not cooperative. Based on these findings, the dependency between the *Blank Signal* anomaly and its causes has been summarised in Figure 2.7.

The main feature for the *Blank Signal* Anomaly is that no ECG features can be observed from the captured. As illustrated in Figure 2.6b, the signal is nearly a horizontal line and no ECG feature is appeared in the plot although there might occasionally be some tiny fluctuation.

Noisy Signal Anomaly

As introduced in Section 2.1.1, the AFE in ECG signal acquisition devices usually has a voltage gain of 500 or more so that the ADC can digitise the signals from the skin. Subsequently, it is extremely sensitive to other signals source as well. The captured signals other than the ECG signal are considered as the noise. In traditional ECG acquisition devices, the AFE may have the ability to handle the noise well. However, the AFE in ECG-based HC-BSNs can not handle the noise well compared to traditional devices. Thus, the most discussed type of usage-related anomaly is the *Noisy Signal* anomaly. In [33], Harrigan et al. summarised
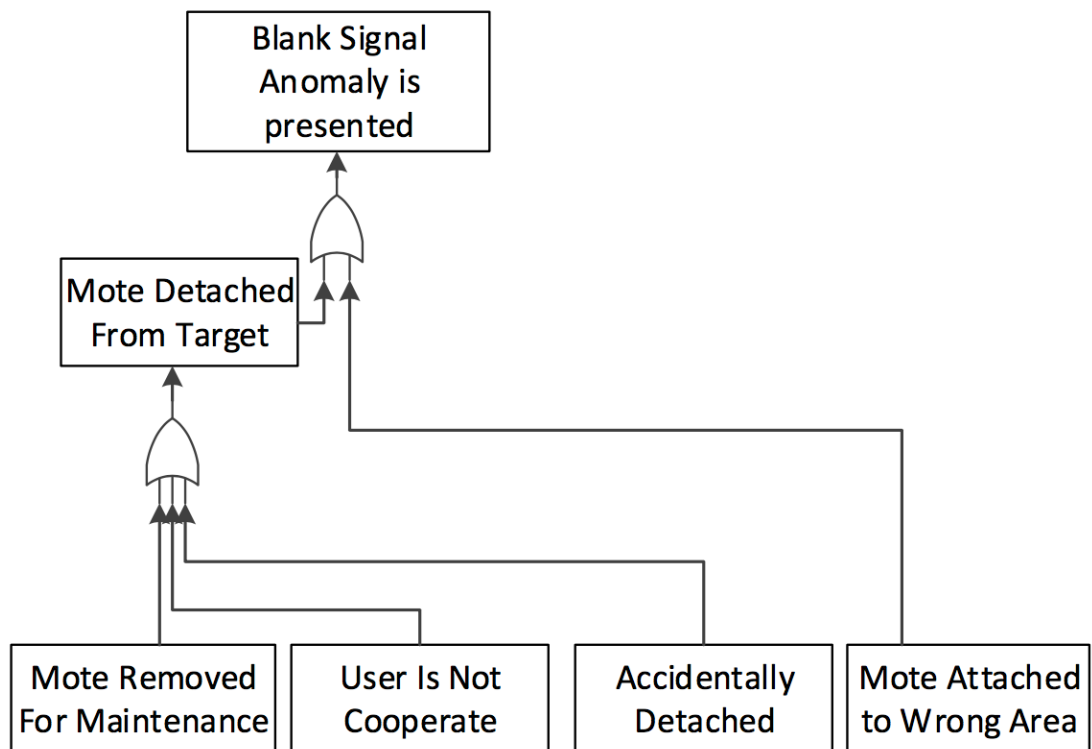
Figure 2.7: The dependency between the *Blank Signal* anomaly and its causes

that the causes to the *Noisy Signal* anomaly can be classified into non-physiologic and physiologic artefact. The non-physiologic artefact is referred to the equipment problem or the environment problem, which is out of the scope of usage-related anomaly. Thus, it will not be discussed in here. The physiologic artefact is referred to the interference on the skin or the muscle activity. According to [25], this type of artefact is generally considered as the most troublesome. In the usage perspective, there are three causes to this type of artefact, subsequently, the *Noisy Signal* anomaly. The first cause is the mote being attached to wrong area so only noise can be captured. Second cause is the monitoring target is performing intensive body movement. The intensive body movement can not only generate muscle activity but also generate static noise as the cloth can rub on the skin. The third cause is the insecure attachment. As the attachment is insecure, the devices can also rub on the skin or cloth to generate more static noise. The insecure attachment can be further broken down into two sub-causes. It can either be the mote is insecure attached or the electrodes including their cable is insecure attached. The

later one leads to the poor isolation of the electrodes or cables. Based on these findings, the dependency between the *Noisy Signal* anomaly and its causes has been summarised in Figure 2.8.

For a signal containing noise, it does not necessarily mean the signal contains *Noisy Signal* anomaly. Depending on the strength of the noise, the ECG signal may be corrupted or may not. If the signal is not corrupted by the noise (as illustrated in Figure 2.6c), the ECG features are still visible. Subsequently, it is categorised as anomaly free. For the corrupted signal (as illustrated in Figure 2.6d), the ECG features are no longer visible. Subsequently, it is declared as containing *Noisy Signal* anomaly. However, The tolerance to the noise can vary between different cardiologist. The main feature of the *Noisy Signal* anomaly is the random pattern in time domain. Figure 2.6d only illustrates a short period of signal. If segmented the signal into several windows, it may seem that the patterns in each window are similar. However, with the length of the signal increase, the number of different patterns for *Noisy Signal* anomaly will increase.

2.2 Literature Survey of Anomaly Detection

The word "Anomaly" is defined in the Oxford Dictionaries as "Something that deviates from what is standard, normal, or expected" [80]. A more technical definition of anomaly is "the patterns in data that do not conform to a well defined notion of normal behavior" [17]. When an anomaly happened, it can result as from the financial lost (e.g credit card fraud) to the lost of life (e.g ECG signal when heart attack). Due to the potentially severe consequences, it is important to detect that an anomaly has occurred within an acceptable duration. Although such jobs can be achieved by a human, it becomes unviable when the scale of the data is large or the distinguishability between normal and abnormal exceeds a normal human's ability. For example, it may not be a difficult job for a human to find a possible credit card fraud from one individual's account. However, once a bank want to prevent the possible fraud to all its customers, such job has exceeded a human's ability as there are thousands or millions of customers in a bank. If such huge
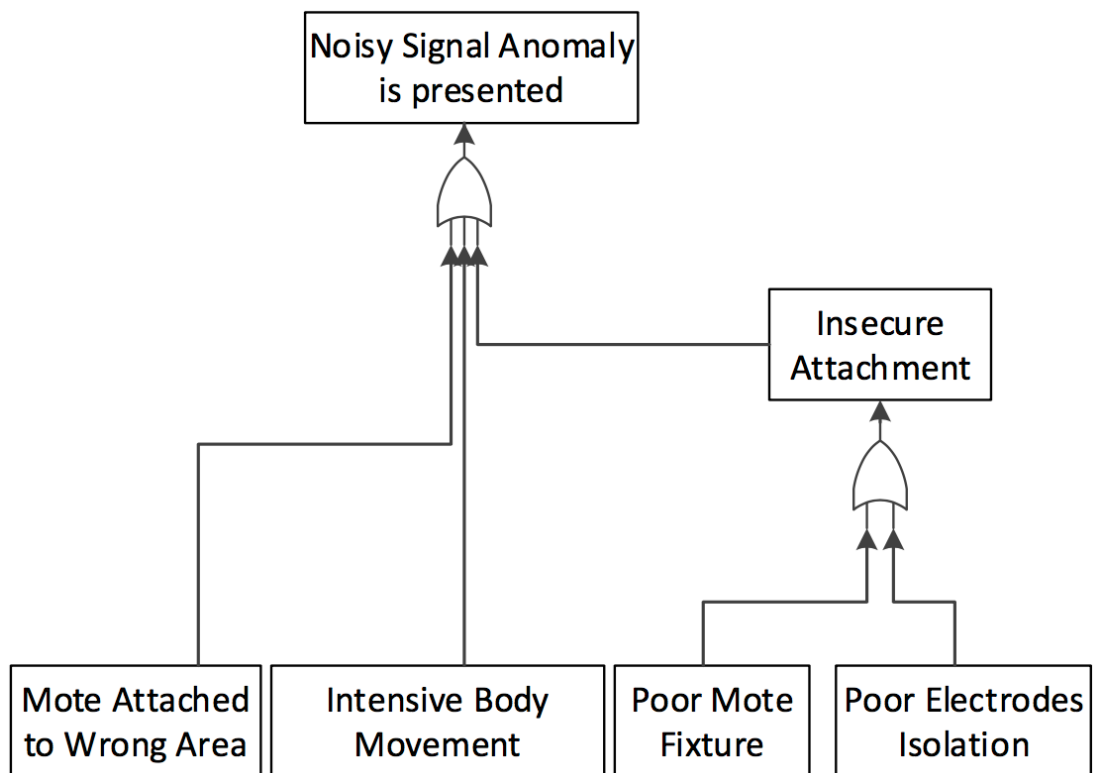
Figure 2.8: The dependency between the *Noisy Signal* anomaly and its causes

job is done by human manually, it is either impossible for one people to do the whole job, or expensive to hire many peoples to do the job. Thus, there are some urgent needs of a mechanism which can detect the anomaly autonomously using computing power. Such mechanisms are usually known as anomaly detection, which has been researched within many research areas and different application domains due to its importance. Chandola et al. defined anomaly detection as "the problem of finding patterns in data that do not conform to expected behavior" [17]. In order to accurately and efficiently perform the anomaly detection, it is important to first clarify which type the targeted anomalies belong to, how the labeled data instances are provided, and which types of approaches can detect the targeted anomaly accurately based on previous two points.

2.2.1   Types of Anomaly

Depending on how a data instance or a sequence of data instances can be considered as an anomaly, there are three types of anomalies: point anomaly, contextual anomaly and collective anomaly.

Point Anomaly

The first type of anomalies is called point anomaly. It is the simplest type of anomaly. A point anomaly means that a single data instance can be considered as an anomaly if that data is deviated from rest of the data instances.

A real-world example of point anomaly will be as following. It is assumed that there is a credit card fraud detection mechanism which classifies a spending as being fraudulent solely based on the spending amount. In this case, if there is one spending of £1,000 whereas all other spendings are no more than £100, the fraud detection mechanism will classify the spending of £1,000 as fraud. As the fraudulent spending is individually existed but significantly deviated in respect to other spending, it belongs to the point anomaly.

Contextual Anomaly

The second type of anomalies is called contextual anomaly. It still describes whether a signal data instance is anomalous or not. Compared to point anomaly, the anomalous data instance is not necessarily deviated from the rest of the dataset. However, the context in which the data instance is appeared is going to classify whether the data instance is anomalous.

A real-world example of contextual anomaly will be the temperature reading of a city. For example, if a sensor reports $-5°C$ at York, it can not be concluded that the reading is anomalous as it is possible for York to get down to $-5°C$. However, if such reading is the temperature in York in August, it can be classified as anomalous temperature reading as it is nearly impossible for York to get down to $-5°C$ in August. In this case, the York and the August are the contexts for the temperature reading.

Depending on the problems that the anomaly detection mechanism are facing, there are two attributes that can define whether a data instance is anomalous:

- **Contextual Attribute:** Contextual attribute describes the context of a data instance. In above example, the month and the location that the temperature is obtained are the contextual attributes as these two attributes regulate the possible normal temperature range which can then be used to decide whether the temperature reading is anomalous.

- **Behavioural Attribute:** Behavioural attribute describes the behaviour that the data instance is representing. In above example, the reading of $-5°C$ is the behaviour that data instance.

For a data instance to be free from anomaly, it is required that both contextual attribute and behavioural attribute to be correct. Either attribute that is anomalous will make the data instance become anomalous. For example, the contextual attribute in above example, which is the month and location, and the behavioural attribute, which is the temperature, must be all correct so that the data instance is free from anomaly. If either the month, the location, or the temperature is anomalous, that data instance will become an anomaly.

Collective Anomaly

The third type of anomalies is called collective anomaly. Compared to point anomaly and contextual anomaly, collective anomaly does not describe whether a single data instance is anomalous or not. Instead, it describes whether a sequence of data instances is anomalous in respect to the rest data set. In [17], collective anomaly is defined as a collection of related data instances is anomalous with respect to the entire data set. When a collective anomaly is appeared, the value of an individual data instance is within the normal range and the location of that data instance is also in a reasonable location. However, the overall pattern of that specific anomalous data sequence is different from the pattern of the correct data sequence. In [17], the authors give an example of a typical web based attack by a remote machine followed by data copying operation from the server to the remote machine via ftp. In that example, a sequence of actions performed by remote machine is like below:

...http-web, buffer-overflow, http-web, smtp-mail, ftp, ssh, http-web, ssh,

buffer-overflow, ftp, http-web, ftp...

In this action log, any of those actions individually is normal. Also, any of these actions happened individually at any time is also normal. However, if a pattern of a sequence actions (buffer-overflow, ssh, ftp) has been logged, that specific pattern may indicate an web based attack followed by data copying anomaly.

2.2.2 Types of Data Labels

A system will need to know how to classify the data between normal and abnormal. The knowledge of normal and abnormal is learned from the existing data labels. The data label means for a given set of data, it tells the algorithm whether the data is normal or abnormal. Depending on how the data label is provided, the anomaly detection algorithm can be categorised into three categories:

Supervised Anomaly Detection

Supervised anomaly detection assumes that the anomaly detection algorithm can learn what is normal pattern and what is abnormal pattern, which are provided by the labelled training data instance. In [17], the supervised anomaly detection is described as to build a predictive model for normal and anomaly data. Then, such model is used to predict whether a new data instance or a new sequence of data instance is normal or anomaly. For a supervised anomaly detection to perform well, it is with paramount importance that the data label has covered all or enough possible normal or anomalous patterns. However, such coverage may sometimes become extremely challenging depending on applications. For example, it is extremely hard to obtain a comprehensive coverage of the possible pattern of static noise. Thus, it may significantly limit the capability of a supervised anomaly detection approach.

Unsupervised Anomaly Detection

In contrast to the supervised anomaly detection approahces, the unsupervised anomaly detection does not require any pre-labelled data [48]. The unsupervised

anomaly detection approach works based on the assumption that within a block of data, the majority data instances are free from anomaly. If a data instance is deviated from the majority data instances to a certain degree, that data instance is considered as anomaly. The unsupervised anomaly detection may become significantly suitable for real world applications (e.g WSNs application) due to the fact that the behaviour of the system is hard to estimate before deployment [48]. However, the key drawback of the unsupervised anomaly detection is that once the assumption of the majority data instance being free from anomaly does not exist, the detection accuracy of this type of approaches will suffer from high false alarm rate [17].

Semi-supervised Anomaly Detection

The third type of anomaly detection approaches is called semi-supervised anomaly detection. The way how this type of apporaches works sits in between the supervised and unsupervised anomaly detection approaches. Most of semi-supervised approaches only requires pre-labelled normal data. Alghough there are some existing works that require only the pre-labelled anomalous data, they are not commonly adapted as it is more difficult to comprehensively cover every possible anomalous data than normal data [17]. This type of detections compares the new data instance to the pre-labelled normal data or the model built on the pre-labelled normal data. If the new data instance is deviated from the pre-labelled normal data, it is considered as anomalous data. This type of approaches are significantly useful for those applications whose normal behaviour can be easily estimated while the anomalous behaviour is hard to estimate. Consequently, the semi-supervised anomaly detections are more widely applicable than supervised techniques [17].

2.2.3   Anomaly Detection for ECG Signals

Following the review of the basic knowledge of anomaly detection, this section will look into the existing anomaly detection approaches which are proposed for ECG signals. First of all, through combining the ECG signal features and the review of anomaly type, it is clear that the anomaly in ECG signal falls into the collective

anomaly. For the ECG signal, the value of each sample does not decide whether the whole signal is anomalous. Also each sample point itself does not contain any contextual information. Whether the signal is anomalous is decided by a sequence of sample points, whose pattern does not comply with what is expected. Subsequently, the anomaly in ECG signal is the collective anomaly. Then, for the features of the usage-related anomalies, the previous survey of ECG has suggested that each usage-anomaly, when happens, will have significant time-domain morphological variation and the ECG features are no longer extractable. Therefore, this review of existing anomaly detection for ECG signals will look into the ability to detect the time-domain morphological variation and whether the detection requires the use of ECG features. On top of this, the review will also look into how the existing works are evaluated including the evaluation setup and the evaluation metrics.

So far, nearly all of existing ECG anomaly detection approaches focus on detecting the anomalous signals which are caused by heart conditions. For the heart conditions, the research in cardiology has already defined the signal features for the known heart conditions. Subsequently, these works mainly focus on detecting anomaly based on different features, and using the different method to extract the desired features. For example, Hadjem et al. [29] has presented an anomaly detection to detect CardioVascular Diseases (CVD). Their approach uses many filters with different cut-off frequencies to remove the noise from the signal first. Then, the QRS-Complex location and duration, T-Wave duration, and P-Wave duration are extracted as the features. They have applied 7 different supervised learning classifiers to the test signals to see which classifier has the best detection accuracy against the CVD. Zheng et al. [93] has also proposed an anomaly detection approach against the CVD. Their approach extract the ST Segment feature and detect the anomalies in ST Segment. In their approach, they use the Support Vector Machine (SVM) as the classifier. There are also some other works have been proposed to detect T-Wave alternans (TWA). These works are also focusing more on the feature extractions. For example, Bortolan et al. [14] proposed an anomaly

detection for the TWA using Principal Component Analysis (PCA) to extract the feature. For these works, as they are designed to detect the anomalous signals caused by heart conditions, the detection is carried out by classifying the features extracted from the signals. These features are defined for medical purpose. However, the main feature of the usage-related anomaly is that the ECG feature is not longer extractable. Subsequently, above works may not be suitable for detecting the usage-related anomalies. However, there are two existing works which are designed to detect the anomalous signal caused by Arrhythmia by comparing the time-domain morphological change. Keogh et al. [39] proposed an algorithm called BFDD , which is an unsupervised algorithm, and in their application case study, they applied BFDD to some ECG signals from the MIT-BIH database [57]. During the operation, It records certain durations of signal. The recorded signal is then segmented into fixed sized windows. Each segment of the signal is then compared to each other by substracting one to another. The segment that has the highest difference to the rest segments is decleared as anomaly. Chuah et al. proposed another anomaly detection algorithm called adaptive window discord detection (AWDD) based on BFDD. Instead of using fixed size window, AWDD use an adaptive window size. As these two approaches compare the segmented signals' time domain difference, they seem to be able to fit for the purpose of detecting the usage-related anomalies.

In terms of the evaluation in these works, the following evaluation setup and evaluation metric are used. To perform the evaluation, signals, which contain the anomalies that are with interests to the algorithms, are fed to the algorithms. Specifically for these works, the signals, which contains anomalies caused by heart conditions, are used. These signals have been pre-labelled for whether it is anomalous or free from anomaly. These labels are used as the ground truth. Then, after the signals are fed into the algorithms, the results, which are reported by the algorithms, can be compared with the ground truth so that the evaluation metrics can be calculated. In these works, the detection accuracy (denoted as DA), which is defined as the ratio between the number of detected anomalies over the total number of

anomalies in the test signals (as shown in Equation 2.1), is used as the evaluation metric. For the test signals, there are many existing signal databases which contain signals recorded from real patients with a specific type of heart conditions. For example, one of the widely used database is called MIT-BIH database [57]. This database contains 48 fully annotated half-hour ECG signals. The signals in this database can be directly used as the test signals and the annotations are used as the ground truth.

$$DA = \frac{Number\ of\ detected\ anomalies}{Total\ number\ of\ anomalies\ in\ test\ signal} \qquad (2.1)$$

Following the above evaluation methodology, the results of existing works are as below. Hadjem et al. have shown that the C4.5 Decision Tree has the best detection accuracy against the CVD. Their approach can achieve 92.54% detection accuracy. Zheng et al. have shown that their approach can improve the detection accuracy to 93%. For the BFDD and AWDD, the detection accuracy achieved is between 40% to 70% and 70% to 90% respectively. Except the detection accuracy achieved by BFDD, other approaches can achieve the desired detection accuracy for this research (over or around 90%). However, as these works are all focus on detecting the anomalous signals caused by heart conditions, none of existing works have performed the evaluation using signals with usage-related anomalies. Subsequently, although existing works can achieve quite high detection accuracy, they may still not be able to detect the usage-related anomalies. Furthermore, there is no existing database with annotation that contains the signals with usage-related anomalies. That means none of existing databases can be directly used to evaluate the detection accuracy of existing approaches against the usage-related anomalies.

2.2.4 Obtaining Evaluation Signals

To overcome the issue of lacking evaluation signals, there are two possible approaches from existing literatures. The first approach is to follow the way of how existing signal databases are created. For example, the MIT-BIH database was

created by Beth Israel Deaconess Medical Center and Massachusetts Institute of Technology through recording the signals from real patients and manually label each RR-Interval by the experts. The main problem with this approach is that it is extremely time consuming and it requires the permission to experiment against participants. The creation of MIT-BIH database lasted for over four years with countless number of participants. For this research, there is no sufficient time as well as participants to create a new database for evaluation. Thus, an alternative solution, which can create the evaluation signal database quickly but also fulfil the evaluation requirement, need to be found, which leads to the second approach.

The second approach to obtain the evaluation signals is to use the fault injection techniques. According to Ziade et al. [94], the fault injection is defined as a "validation technique of the dependability of a system which consists in the accomplishment of controlled experiments where the observation of the system's behavior in presence of faults is induced explicitly by the writing introduction (injection) of faults in the system." The aim of using fault injection is to determine whether a system will response in a desired way when certain faults are purposely introduced into the system. Specifically to this research, the aim will be to evaluate whether the algorithm can report a positive result when a certain usage-related anomaly is purposely introduced into the ECG signal. So far, there is only one but widely acknowledged and used work that use such approach to construct the evaluation signals. In [58], Moody et al. created a noise stress test database based on MIT-BIH database using fault injection approach. This database aims at evaluating the arrhythmia detectors' accuracy when the ECG signals containing noise with different noise strength. Their approach is to record the noise from participants first. Then, the recorded noise is introduced into the ECG signals from MIT-BIH database with the noise strength being controlled. The introduction of noise is basically adding the noise to the ECG signals in a sample-to-sample basis. As their aim was to evaluate the arrhythmia detectors' accuracy when noise is presented in the ECG signals, they still used the original annotation in the MIT-BIH database. As Moody's approach is widely acknowledged and used in the researches related

to ECG signals, it is believed that it can also be used in this research. However, several modifications will need to be made in this research's methodology due to the different research purpose. A possible modification will be the fault injection consistency. For example, Moody's work injected the noise to the ECG signals from the beginning to the end of the ECG signals. In contrast, this research may require the noise being injected with a random length so that the evaluation signal is a mixture of anomaly-free RR-Interval and anomalous RR-Interval. By doing this, it can better evaluate the proposed algorithm's ability to detect usage-related anomalies as well as to avoid false result being reported.

2.3 Summary

In this chapter, the literature survey of the ECG and the anomaly detection has been presented. In the perspective of ECG, the review has summarised the different types of ECG devices. Then, the deployment of the ECG device, which is used in this research, has been reviewed. How the ECG signal is composed of when no usage-related anomaly is presented has also been reviewed. Then, the dependencies the usage-related anomalies and their root causes have been defined based on existing literatures. For the anomaly detection, this chapter has reviewed the basic knowledge to anomaly detection. Then, the existing anomaly detection approaches for ECG signals have also been reviewed. Through this review, it is clear that no evidence has been shown by existing works that they can fit for the purpose of this research. Subsequently, next chapter will look into proposing a new anomaly detection methodology.

Chapter 3

USAGE-RELATED ANOMALOUS SIGNALS DETECTION

This chapter will focus on detecting the usage-related anomalous signals in ECG-based HC-BSNs. To better address the problem, this chapter starts with listing the objectives (Section 3.1) that are desired to achieve followed by the research challenges towards those objectives (Section 3.2). Then, in Section 3.3, an anomaly detection algorithm named as *AID* will be proposed. It aims at detecting the previously concluded usage-related anomalous signal. The proposed algorithm will be evaluated using Matlab simulation and on-mote simulation. The evaluation will be presented in Section 3.4 and Section 3.5 respectively. Finally, this chapter will be summarised in Section 3.6.

3.1 Objectives

Following the idea of assisting the user to use the ECG-based HC-BSNs correctly by notifying the user when the usage becomes erroneous, there are three core objectives for the desired anomaly detection algorithm:

Perform Detection Accurately

The detection accuracy is the fundamental requirement. An inaccurate detection algorithm, no matter it is with high false-positive rate (erroneous usage is reported when no usage error is made) or with high false-negative rate (erroneous usage is not reported), will lead to the usage issues being untreated, or the user is annoyed by the system. In both cases, the users will lost the trust to the algorithm. Subsequently, the assistant from the system will be ignored by the users and the usage issues may not be addressed.

Perform Detection At Run-Time

The ECG signal is periodic and each cycle of the signal reflects the condition of each heartbeat. The heart disease diagnosis from the ECG signal is not only depending on the parameters extracted from many ECG cycles (e.g heart-rate), but also depending on the pattern of each cycle of the ECG signal. As a consequence, it is important to guarantee that each cycle of the ECG signal is correct, and the signal is continuously correct. Subsequently, if the proposed algorithm can be run at run-time, it would be capable to guarantee that each cycle is correct, or notify the user of erroneous usage without any delay. On the other hand, if the algorithm can not run at run-time, that means the system will collect a batch of data before making sure the signal is correct. In this case, the whole batch of data will be with no useful information if the usage error is made at the initial deployment phase. Or, the batch of data will be partially useless if the usage become erroneous during the normal operation phase of the system. In both case, the notification of erroneous usage to the user will be delayed and the correctness of the data is not guaranteed. Furthermore, the system is wasting energy to collect and transmit useless data while the patients are in danger due to the delayed notification.

Perform Detection On A Mote

As the algorithm is required to be run at run-time, it needs to be able to run on a mote. Otherwise, the transmission of collected data is required, which will lead to

Mode	mica2	mica2dot	micaz
Rx	9 mA	9 mA	18.8 mA
Tx (0dBm)	15 mA	15 mA	17.4 mA
Power Down	10^{-3} mA	10^{-3} mA	10^{-2} mA

Figure 3.1: The current drawn from the battery for different radio operating modes for some widely used motes (quoted from [11]). It can be seen that when the radio is turned off, the current drawn from the battery is only around $10^{-3} mA$ to $10^{-2} mA$. In comparison, when the radio is turned on, no matter it is in transmission mode or receive mode, it will draw between a hundred to a thousand times more current from the battery.

communication overhead. For a ECG-based HC-BSN system, the sampling frequency will be from several hundreds Hz to thousands Hz while the sampling resolution will be from 8-bit to 24-bit. Using the parameters ($512Hz$ sampling frequency with 12-bit sampling resolution) from the devices (Shimmer2r [2]) used in this research as an example, for every one second of signal, it will generate 6 kb data. Such data rate may not sound huge comparing to other scenarios (e.g internet browsing or video streaming). However, considering the wireless communication will cost many magnitudes more energy than computation [11] (the current drawn from the battery for different radio operating modes for some widely used motes is demonstrated in Figure 3.1), and the battery on the mote is low volume (e.g battery on Shimmer2r is only 450 mAh), the battery will be drained rapidly if the running of the algorithm requires frequent wireless communication. For example, an experiment of using the Shimmer2r to live stream ECG signal to laptop via Bluetooth has shown that the on mote fully charged battery will be drained in around one hour. On the other hand, if the radio is turned off, but the mote is sensing signal and writing the signal to SD-Card, the fully charged battery can last for several days. Based on these findings, it is desired that the algorithm has the capability to perform the detection on a mote.

3.2 Research Challenges

In order to achieve the desired objectives that have been listed in Section 3.1, the research challenge will need to be clarified first so that when the algorithm is

designed, the challenges can be addressed. In this section, the possible research challenges will be discussed and explained. As the algorithm is desired to be run on a mote at run-time, there are some obstructions which is inherent from the mote platform. The following research challenges are listed as below:

- **Limited On-board Memory -** For a typical BSN mote, the on-board memory is usually constrained. Using the Shimmer2r as an example, it has 10 Kbyte of RAM and 48 Kbyte of ROM. Other platforms like smartphone can have hundreds of Mbyte or even many Gbyte of RAM with many Gbyte of ROM. The problem is made worse as there are other software components (e.g network stack, OS Kernel etc.) that have demand from the memory. As a consequence, in order for the algorithm to run on a mote, the memory overhead need to be kept within an acceptable level.

- **Limited Computational Ability -** In order to preserve the energy from the on-board battery, a BSN mote usually equips with a microcontroller (MCU) with limited computational ability. Again, using the Shimmer2r as an example, there is a MSP430F1611 MCU on-board. This MCU integrates a 16-bit processor running at up to 8 MHz. These parameters mean that the computational ability of the on-board MCU is extremely constrained compared to other processors that are found in other platforms. For example, the processors in nowadays' smartphone, laptop and desktop are 32-bit or 64-bit running at several GHz. Due to the constrained computational ability, the algorithm's complexity is limited if it runs on a mote.

- **No Floating Point Number Support -** Floating point number processing is neither supported by the hardware, nor the software. The MCU on the Shimmer2r, which is MSP430F1611, does not equip with the hardware floating point unit. Meanwhile, the compiler of Contiki OS also does not support the software implementation of the floating point number processing. As a consequence, it significantly limits the algorithm's ability of extracting features from other domain. For example, it is quite common that the frequency do-

main feature is extracted and used for detecting the anomaly. The conversion between time domain and frequency domain requires performing the Fourier Transform, which heavily relys on the capability of floating point number processing. If the floating point number is not supported, the domain conversion can be hardly performed. Subsequently, the frequency domain feature is not available for anomaly detection.

- **Timing Requirement -** As the algorithm need to be run at run-time, an algorithm must be able to perform the detection within a certain time window. Depending on the frequency that the detection is performed, the window may vary. For example, if the detection is carried out for each RR-Interval, that means the time window is the length between two R-Peaks. The algorithm starts the detection after current R-Peak and must finish the detection before the next R-Peak arrived. Considering the normal heart-rate is between 50 bpm to 100 bpm [38], the deadline for performing one detection is decided by the highest heart-rate, which is 0.4 second.

- **Limited Data Label -** As being introduced in Section 2.1.4, how to justify the correctness of an ECG signal has been well defined by cardiologists so far. The patterns of ECG signals with certain types of heart diseases have also been defined by the cardiologists. However, as this thesis looks into the anomalous ECG signals caused by the erroneous usages, the patterns of these types of ECG signals can vary in a wide range. For example, the pattern of the static noise signal, which can be captured by the ECG devices, does not come with a specific pattern. As a consequence, the data label may be hard to comprehensively cover the whole range of possible signal variation.

3.3 *AID*: Lightweight Usage Anomaly Detection

Through the review of existing anomaly detection approaches, it is clear that none of existing ECG anomaly detection approaches will work for detecting the usage anomalies on a mote and at run-time. As a consequence, it is necessary to propose

an anomaly detection algorithm which is capable of detecting the usage-related anomalies so that the users can be notified when they use the ECG devices erroneously. In this section, an anomaly detection algorithm called *AID* is presented followed by a scenario-based assessment for the proposed algorithm. Finally, how the proposed algorithm is implemented on a mote will be introduced in order to show how the limitations (e.g no floating-point number support on a mote) is addressed.

3.3.1 Algorithm Design

To achieve the desired objectives, in this section, the high level design decisions are made and will be discussed. There are three main design decisions that have been made for *AID*:

- Performing the anomaly detection against each RR-Interval

- Adopting the semi-supervised detection approach

- Performing the anomaly detection in time domain by comparing two time series data sequence

In the rest of this section, each of these decision will be justified and the key issues that will need to be addressed by the algorithm will be discussed.

An ECG signal is periodic, and the length of each period is depending on the heart-rate. To extract the heart-rate from an ECG signal, measuring the length of the interval between two adjacent R-Peaks is the common approach. In other word, an ECG signal is repeated between two R-Peaks. Furthermore, as the R-Peak is the most significant feature from an ECG signal, it is easier to detect a R-Peak compared to other ECG components. Easy to detect potentially means less complex algorithm will need to be applied. Subsequently, it is more likely to be run on a mote at run-time. Due to these two reasons, *AID* is designed to perform the anomaly detection against each RR-Interval. The detail of how *AID* detects the R-Peak will be covered in Section 3.3.2.

As introduced in Section 3.2, there is limited data label for the anomalous signals caused by erroneous usages. On the other hand, it is well acknowledged about in which form an ECG signal is correct without usage anomaly. Due to these two points, *AID* is designed to adopt the semi-supvervised detection approach. That means, the *AID* compares the new RR-Interval to the correct RR-Interval, and other than the correct RR-Interval, *AID* classifies as the anomaly. The correct RR-Interval is set by cardiologist during the initial deployment of the system. The detail of the initial deployment of *AID* will be covered in Section 3.3.5.

The third design decision is to perform the detection in time domain. As it is desired to run on a mote at run-time, the obstructions of constrained on-mote environment will significantly limit the ability to run any complex algorithm. The transformation between time domain and frequency domain is normally done using Fourier Transform. Although there is Fast Fourier Transform algorithm which can accelerate the calculation, it still requires $N * log_2 N$ floating point multiplication operations, in which the N is the number of data points which depends on the sampling frequency and the frequency domain resolution. Assuming the sampling frequency is $256Hz$ and the frequency domain resolution requirement is $1Hz$, then, the N is 256. As introduced in Section 3.2, there is limited computational ability and no floating point number support. As a consequence, the domain transformation should be avoided. Subsequently, *AID* is designed to perform the detection in time domain.

To perform the detection in time domain, there are three important issues that will need to be addressed in order to assure the detection accuracy. The first issue is the signal phase difference. In this thesis, overcoming the signal phase difference is referred to achieving phase invariance. Figure 3.2 illustrates an example of the issue of phase difference between two signals. In the figure, it shows two RR-Intervals represented by blue line and red line respectively. In terms of the correctness of these two RR-Interval, they are anomaly free. However, directly comparing these two signals in time domain will result as one being not corresponding to another. In other word, if it is assumed that the red signal is the correct RR-Interval and the blue signal is the new RR-Interval, the anomaly detection should

give negative result as the blue signal is anomaly free. However, the anomaly detection will give positive result as the two signals are not aligned in time domain, which means the blue signal does not comply with the red signal in a point by point basis. This example shows the importance to achieve phase invariance. In *AID*, phase invariance is achieved using R-Peak alignment approach. That means two signals are aligned using the detected R-Peak, which is labelled in Figure 3.2 by the '*'. The result of R-Peak alignment is shown in Figure 3.3.
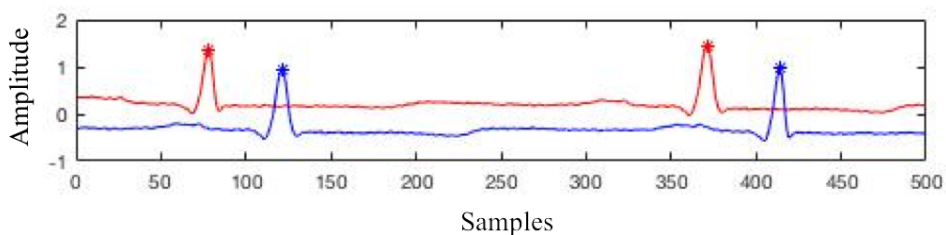
Figure 3.2: An illustration of the phase difference between two signals
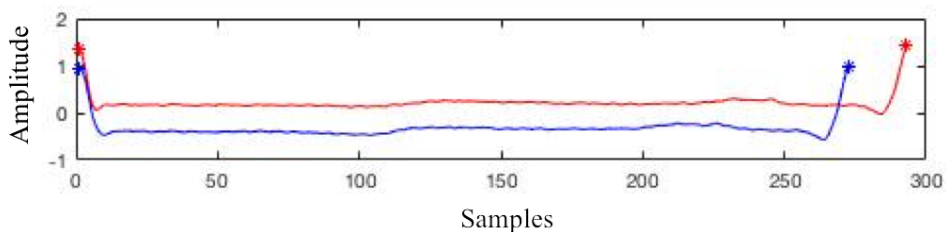
Figure 3.3: An illustration of the R-Peak alignment and the size difference between two signals

After two signals are aligned, the second issue is the signal length difference. In this thesis, overcoming the signal length difference is referred to achieving size invariance. As illustrated in Figure 3.3, two signals (red and blue) are now aligned by the first detected R-Peak of each signal. However, from the figure, it is clear that two signals are different in size. The red signal is around 30 samples longer than the blue signal. In this case, two signals can not be compared in time domain due to the size difference. This example shows the importance of achieving size invariant. To achieve size invariant, *AID* uses a mechanism called RR-Interval Prediction. The detail of how the RR-Interval Prediction mechanism works will be
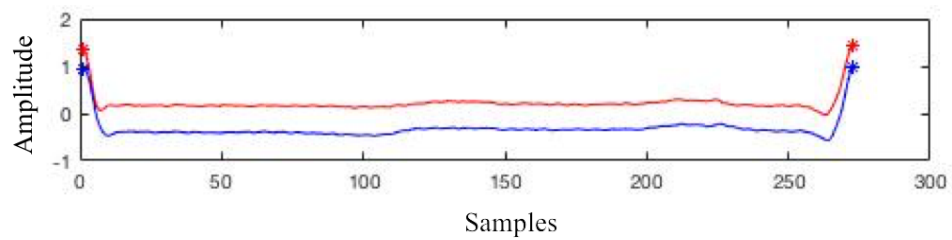
Figure 3.4: An illustration of achieving the size invariant by *AID*

covered in Section 3.3.3. In short, the RR-Interval Prediction mechanism extends or shrinks the non-cardio-activity duration of a RR-Interval so that the size of the RR-Interval is modified without losing any detail of those ECG components. The result of the RR-Interval Prediction is illustrated in Figure 3.4.

The third and the last issue that need to be addressed is the difference in two signals' amplitude. In this thesis, overcoming the signal amplitude difference is referred to achieving amplitude invariance. As shown in Figure 3.4, the two signals are now aligned and size matched but there is still difference in amplitude. Such amplitude difference may happen in but not limited to the following situations. First, the capacitance in human skin may lead to the different base voltage on the skin. The ECG signal is varied on top of the base skin. Second, the raw Analogue-to-Digital Converter (ADC) reading may vary due to the reference voltage change depending on the hardware implementation. Third, the strength of ECG signal from an individual may also vary during different phases of a day (e.g ECG signal may be weaker when an individual is sleep compared to when the individual is excited). The ECG signal can be calibrated according to some parameters (e.g baseline voltage, ADC gain error etc.) which are used when the signal is being captured and digitised. The signal calibration itself is basically shift and scale the raw signals based on those parameters. However, to extract these parameters, it will require more complex algorithm to pre-process the captured signals as well as some extra hardware or manual operations. For example, the calibration process described in the Shimmer ECG User Guide [68] requires a signal generator with output voltage being controlled to measure the voltage gain. Then, it requires the ADC readings when no interference and signal are being captured to calculate the baseline volt-

age and ADC gain error. At last, the raw ECG signal is scaled and shifted according to these parameters so that the signal is calibrated. For the application scenario of this research, these parameters are hardly available for the algorithm. Thus, it is desired that the algorithm can achieve the amplitude invariance without the use of those parameters. Therefore, *AID* adopts a statistical comparison approach named Pearson's Correlation Coefficient. As shown in Equation 3.1, the Pearson's Correlation Coefficient, r, is the covariance between two input signals, x and y, divided by the product of the standard deviations of x and y. When calculating the covariance (the numerator in Equation 3.1), it calculates the difference between the data point and the mean value of that data sequence. As a result, the amplitude difference is normalised. Subsequently, the amplitude invariance is achieved.

$$r = \frac{\sum_{i=1}^{n}((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.1}$$

where:

x_i: The ith data point of input data x

$\bar{x}$: Mean amplitude of the input data x

y_i: The ith data point of input data y

$\bar{y}$: Mean amplitude of the input data y

r: Correlation Coefficient

n: The number of samples of the two inputs, x and y, which must be in the same size

In summary to the algorithm design, *AID* is designed to perform the anomaly detection in time domain adopting semi-supervised detection approach. It detects the anomaly by comparing time-domain pattern against the ground truth which is set by cardiologist during the initial deployment (detail covered in Section 3.3.5). During the run-time, *AID* follows the three-stage detection approach (the pseudocode is shown in Algorithm 1:

- **Stage 1 -** R-Peak detection and alignment to achieve phase invariance (the details are covered in Section 3.3.2)

- **Stage 2 -** RR-Interval Prediction to achieve size invariance (the details are covered in Section 3.3.3)

- **Stage 3 -** Signal comparison and classification using Pearson's Correlation Coefficient to achieve amplitude invariance (the details are covered in Section 3.3.4)

3.3.2 R-Peak Detection

There are plenty of algorithms (e.g Pan-Tompkins [63]), which can reliably detect the R-Peak under various condition, available. These algorithms require either domain transform or filters, which are very complex operation and won't fit on a mote. Consequently, the existing approaches may not be suitable due to the complexity.
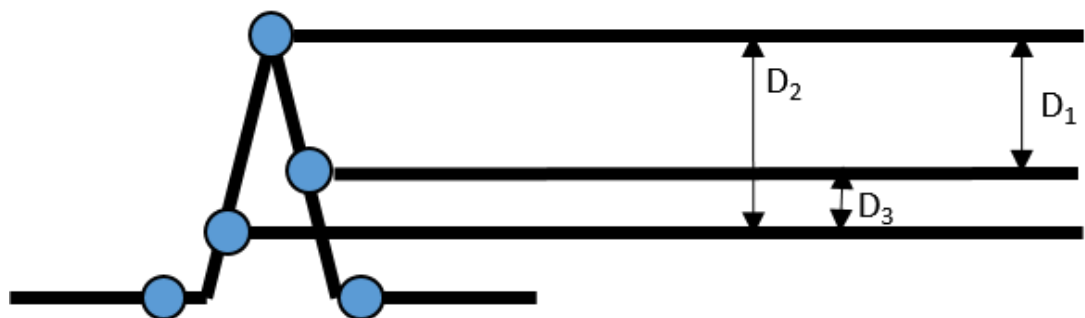
Figure 3.5: The notations of the amplitude difference used to explain the R-Peak Detection in *AID*

In *AID*, a simple algorithm is adopted to detect the R-Peak (pseudocode of R-Peak detection is shown in Algorithm 1 between Line 9 to Line 23). In order to clearly explain the R-Peak detection algorithm, the following notations are used to represent the amplitude difference between two data samples. For each data sample, the absolute amplitude difference between itself and its next data sample is denoted as D_1; the absolute amplitude difference between itself and its previous data sample is denoted as D_2; and the absolute amplitude difference between its previous

data sample and its next data sample is denoted as D_3. These notations are illustrated in Figure 3.5. To perform the R-Peak detection, the algorithm also requires a threshold, denoted as T_R, which is set by cardiologist during initial deployment (covered in Section 3.3.5). During the R-Peak detection, for each data sample, if D_1 and D_2 exceed T_R, and D_3 is smaller than T_R, this specific data sample is reported as the R-Peak. Otherwise, the R-Peak is not detected and the algorithm will perform R-Peak detection for the next data sample until the R-Peak is detected.
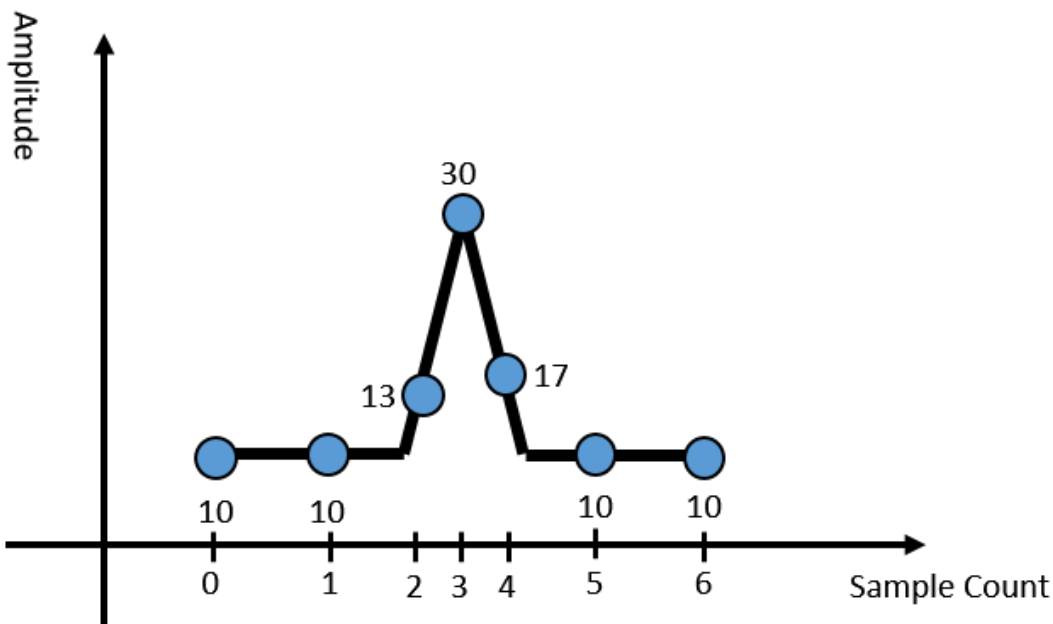
Figure 3.6: An example to explain the R-Peak Detection in *AID*

To demonstrate how the algorithm work, Figure 3.6 has illustrated an example signal with the horizontal axis is the sample count and the vertical axis is the amplitude. Each blue point in the figure represents a data sample whose amplitude is labelled next to the point. For this example, it is assumed that the T_R is set to 10, and the fourth data sample (sample count 3 as count starts from 0) is the R-Peak. When the algorithm check whether the fourth data sample is the R-Peak, it calculates the D_2, which is 17 in this case, and the D_1, which is 13 in this case. Then, the algorithm calculates the D_3, which is 4 in this case. As the D_1 and D_2 is greater than T_R and the D_3 is smaller than T_R, the algorithm reports this point as the R-Peak. On the other hand, let's use the third data sample (sample count 2) to

Page 50

be the example to demonstrate how the algorithm will behave when a data sample is not R-Peak. When the algorithm check whether the third data sample is the R-Peak, it calculates the D_1, which is 17 in this case. Then, the algorithm calculates the D_2, which is 3 in this case. As the D_2 is smaller than T_R, the algorithm does not report this point as the R-Peak.

By using this approach, the complexity is reduced compared to other existing approaches. However, as such approach is proposed to be used with the signals captured from the ECG-based HC-BSNs, it assumes that the sampling frequency of the signals is relatively low. Specifically for this research, the sampling frequency is assumed to be $360Hz$ as the evaluation signal database used in this research is sampled at $360Hz$. Despite this assumption, the proposed R-Peak detection algorithm may be generalised to other sampling frequencies with some modifications. The main modification is the data samples used to calculate the amplitude differences (D_1, D_2 and D_3). For example, if the sampling frequency is $1024Hz$, instead of calculate those amplitude differences between three adjacent data samples (input[Cnt_i], input[$Cnt_i - 1$], input[$Cnt_i - 2$] if using the notations in Algorithm 1), the algorithm can be modified to calculate those amplitude differences using three data samples which are more separated (e.g input[Cnt_i], input[$Cnt_i - 5$], input[$Cnt_i - 10$] if using the notations in Algorithm 1). This modification is based on the fact that with higher sampling frequency, the captured signal will have more data samples than the signal with lower sampling frequency. As the algorithm still calculates the D_1, D_2 and D_3 using three data samples, the complexity in terms of computation will not increase. However, as the algorithm need to buffer more data samples before the R-Peak detection can be performed, it may increase the overhead of RAM consumption. As generalising the proposed R-Peak detection algorithm is outside the scope of this research, how far the data samples should be separated will not be covered in this research.

Referring to the cases that *AID* concerns about, when the captured signal is normal, it can detect the R-Peak reliably as illustrate in Figure 3.6. When the *No Signal* anomaly happens, no R-Peak is exist. As a result, it's expected that the algorithm

doesn't report any R-Peak being detected. Subsequently, the *No Signal* anomaly can be declared. When the *Inverted Signal* anomaly happens, the highest peak become the lowest peak. As our algorithm compares the absolute difference, it will either report the highest peak (correct R-Peak) or lowest peak (inverted R-Peak) as long as they are significant. As a result, the algorithm will work under this situation. When the *Noisy Signal* anomaly happens, depending on the strength of the noise, the algorithm behaves differently. If the noise is much weaker than the R-Peak, the peaks caused by the noise won't affect the detection due to the thresholds, T_R. On the other hand, if the noise is similar or stronger than the R-Peak, the peaks caused by the noise may be reported as R-Peak by this algorithm, in which case the algorithm reports fake R-Peak. However, it's won't affect the *AID* as the R-Peak unrecognisable is one of the key feature of *Noisy Signal*.

3.3.3 RR-Interval Prediction

In order to better explain how RR-Interval prediction algorithm works, the following terms are defined:

- **Correct RR-Interval -** The ground truth that the semi-supervised detection adopted. It is set up by cardiologist during the initial deployment

- **New RR-Interval -** The newly captured RR-Interval during the run-time. Awaiting for being check for usage anomaly

- **Predicted RR-Interval -** The output of RR-Interval Prediction algorithm. Generated from the *Correct RR-Interval* and is in the same length as the *New RR-Interval*. It is the prediction of how the *New RR-Interval* should look like

The objective of the RR-Interval prediction algorithm is to match the signals' length between the **Correct RR-Interval** and the **New RR-Interval**. Some of existing works match the signal length by either using part of the signal with fixed window size (e.g BFDD [39]) or sub-sampling the whole signal (e.g AWDD [18]). These two approaches can, in certain cases, damage the features in the ECG signal.
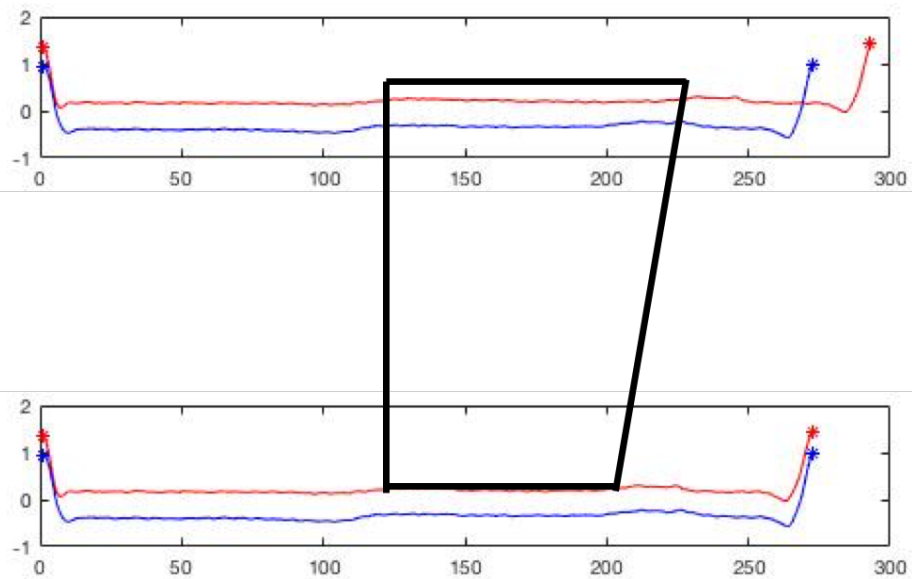
Figure 3.7: An example the RR-Interval Prediction

In *AID*, the signals' length is matched without modifying the ECG components within each RR-Interval. It is achieved through modifying the non-cardio-activity duration in each RR-Interval, which does not contain any ECG components. The pseudocode of RR-Interval prediction is shown in Algorithm 1 between Line 24 and Line 31. For each *New RR-Interval*, the algorithm measures its length. When the *Correct RR-Interval* is shorter than the *New RR-Interval*, the algorithm will perform linear interpolation to the non-cardio-activity duration of the correct signal until the whole signal matches the size of captured signal. On the other hand, if the *Correct RR-Interval* is longer than the *New RR-Interval*, the algorithm will shrink the correct signal by removing the samples during non-cardio-activity duration. The modified signal is then used as the *Predicted RR-Interval*.

Figure 3.7 illustrates an example of the *Correct RR-Interval* being longer than the *New RR-Interval* and how the RR-Interval prediction algorithm matches two signals' length. The red signal in the upper figure is the *Correct RR-Interval* whereas the blue is the *New RR-Interval*. In the lower figure, the red signal is the *Predicted RR-Interval* and the blue signal is still the *New RR-Interval*. The R-Peaks of each signals are labelled by '*'. In this case, as the *Correct RR-Interval* is longer than the *New RR-Interval* in the upper figure, the samples in the non-cardio-activity dura-

tion will need to be deleted correspondingly to obtain the *Predicted RR-Interval* (as denoted by the box with thick black line). The outcome is the *Predicted RR-Interval* in the lower figure in red is with the same length to the *New RR-Interval*.

3.3.4   Signal Comparison and Classification

As introduced in Section 3.3.1, *AID* uses Pearson Correlation Coefficient to compare the similarity between *Predicted RR-Interval* and the *New RR-Interval*. The outcome from Pearson's Correlation Coefficient is an index which is between -1 and $+1$. The $+1$ index means for a given x being increase or decrease, another variable, y, is increase or decrease correspondingly. As illustrated in Figure 3.8a, when x is increased, y is also increased. It is called that x and y are positively correlated. On the other hand, the -1 index means for a given x being increase or decrease, another variable, y is decrease or increase respectively. As illustrated in Figure 3.8b, when x is increased, y is decreased. It is called that x and y are negatively correlated. When the index is 0, that means there is no relationship between the variation of x and the variation of y (as shown in Figure 3.8c). That is known for variable x having no correlation to variable y.

To classify the signal into different types of anomaly, *AID* uses a threshold approach to minimise the computation overhead. The pseudocode of the signal comparison and classification using Pearson's correlation coefficient and threshold is shown in Algorithm 1 between Line 32 and Line 40. The classification threshold, denoted as T_C, is set by cardiologist during the initial deployment. By changing the threshold, T_C, an individual cardiologist can adjust the strictness of the anomaly detection according to their ability of interpreting the ECG signal. When the *New RR-Interval* is in an acceptable range to the cardiologist, it should be correlated to the *Predicted RR-Interval* to a certain degree. In *AID* if the correlation coefficient, denoted as r, is greater than T_C, the *New RR-Interval* is classified as free from anomaly. On the other hand, if the r is smaller than the $-T_C$, that means the *New RR-Interval* is negatively correlated to *predicted RR-Interval*. Consequently, the *New RR-Interval* is classified as *Inverted Signal* anomaly. If the r is between T_C and $-T_C$, that means the *New RR-Interval* is less or not correlated to the *Predicted RR-Interval*.
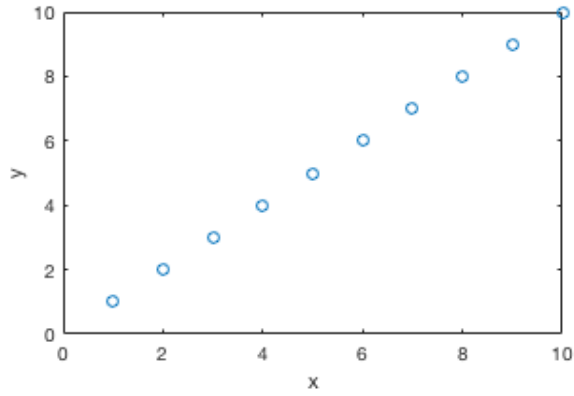
(a) Variable x and y are positively correlated
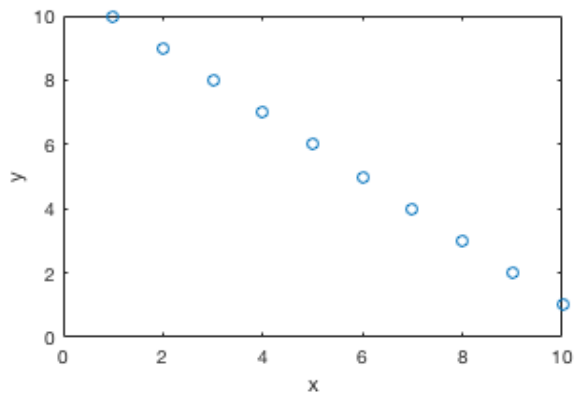
(b) Variable x and y are negatively correlated
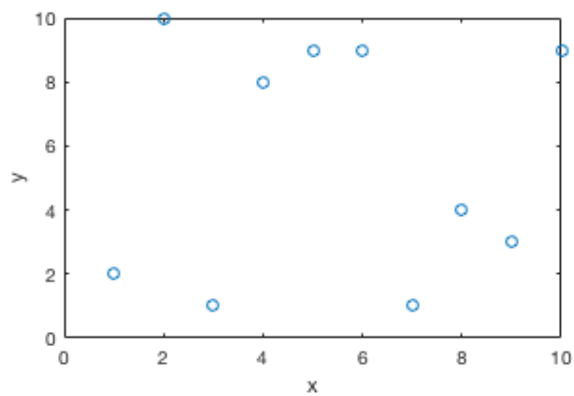
(c) Variable x and y has no correlation

Figure 3.8: An illustration of variable x and y are positively correlated (a), negatively correlated (b) and no correlation (c)

In this case, the *New RR-Interval* is classified as *Noisy Signal* anomaly.

3.3.5 Deployment of the *AID*

AID needs to know what is the correct signal from this specific signal source, what level of noise is accepted by cardiologist, and the minimum heart rate (Maximum RR-Interval length) that is acceptable by cardiologist. During the initial deployment, cardiologist can check the quality of the captured signal until the signal quality meets their requirement. A RR-Interval, which meets cardiologist's best quality requirement (e.g no noise, fully correct ECG signal), will be picked and stored on the mote as the *Correct RR-Interval*. *AID* can then measure the amplitude of the R-Peak and its adjacent samples of the *Correct RR-Interval* to set-up the threshold, T_R, for R-Peak detection. The control of acceptable level of noise is done by setting up the correlation coefficient threshold T_c. The cardiologist can manipulate the deployment (artificially creating noise to the signal) until the signal quality meets their lowest requirement (the worst signal that a cardiologist can interpret). A RR-Interval that meets the cardiologist's lowest requirement will be recorded. Then, the correlation coefficient against the *Correct RR-Interval* is calculated by *AID* and used as the threshold T_c for future detection. The minimum heart rate, which can also known as the maximum RR-Interval length and denoted as RR_{MAX}, is set by cardiologist according to the medical requirements of each patient. Then, the minimum heart rate will be converted to RR_{MAX} depending on the sampling frequency, which will then be stored on the mote. For example, if the cardiologist believes that the minimum heart rate for a specific patient is 30 beat-per-minute (bpm), the cardiologist can enter 30 bpm to the system. Assuming the sampling frequency is $360Hz$, 30 bpm is then equal to each RR-Interval being 2 seconds. Thus, the RR_{MAX} is equal to 720, which is calculated by using 2 second to multiply by $360Hz$. This value is then stored on the mote.

3.3.6 Scenario-based Assessment

In order to demonstrate the effectiveness of the proposed *AID* against the usage anomalies, a scenario-based assessment will be presented in this section. According to the usage anomalies, there are five scenarios that are of interest to the

Algorithm 1 Pseudocode of AID

1: Pre-set parameters: ,
2: T_c: Classification threshold
3: T_R: Threshold for R-Peak detection
4: RR_{MAX}: Maximum RR-Interval length
5: *Correct RR-Interval* RR_C: Correct signal set by cardiologist
6: *Correct RR-Interval* length L_C: The number of samples of the *Correct RR-Interval*
7: *Input*: A new input signal sequence started from previous detected R-Peak
8: L_i: The number of samples of the input signal sequence
   /* Stage 1 - R-Peak Detection */
9: Set Cnt_i to 2
10: **if** $|$ input[Cnt_i-1] - input[Cnt_i-2] $| >= T_R$ AND $|$ input[Cnt_i-1] - input[Cnt_i] $| >= T_R$
    AND $|$ input[Cnt_i-2] - input[Cnt_i] $| < T_R$ **then**
11:     input[Cnt_i-1] is R-Peak
12:     GOTO Stage 2 - RR-Interval Prediction
13: **end if**
14: **while** *input[Cnt_i-1]* is not R-Peak **do**
15:     Cnt_i++
16:     **if** $|$ input[Cnt_i-1] - input[Cnt_i-2] $| >= T_R$ AND $|$ input[Cnt_i-1] - input[Cnt_i] $| >= T_R$
        AND $|$ input[Cnt_i-2] - input[Cnt_i] $| < T_R$ **then**
17:         input[Cnt_i-1] is R-Peak
18:         GOTO Stage 2 - RR-Interval Prediction
19:     **end if**
20:     **if** $Cnt_i >= RR_{MAX}$ **then**
21:         **return** *Blank Signal* detected
22:     **end if**
23: **end while**
    /* Stage 2 - RR-Interval Prediction */
24: size difference = $L_C - (Cnt_i - 1)$
25: **if** *size difference* > 0 **then**
26:     Shrink the non-cardio-activity part of RR_C by removing samples to match Cnt_i-1,
        and use it as *Predicted Signal*
27: **else if** *size difference* < 0 **then**
28:     Perform linear interpolation to the non-cardio-activity part of RR_C until the whole
        signal's size matches Cnt_i-1, and use it as *Predicted Signal*
29: **else**
30:     RR_C is used as *Predicted Signal*
31: **end if**
    /* Stage 3 - Signal Comparison and Classification */
32: Calculate the correlation coefficient between *Predicted Signal* and *Input*
33: **if** correlation coefficient $>= T_c$ **then**
34:     *Correct RR-Interval = input*
35:     **return** *No anomaly detected*
36: **else if** correlation coefficient $<= -T_c$ **then**
37:     **return** *Inverted Signal* detected
38: **else**
39:     **return** *Noisy Signal* detected
40: **end if**

---

author:

- The *New RR-Interval* is free from anomaly

- The *New RR-Interval* contains *No signal* anomaly

- The *New RR-Interval* contains *Inverted Signal* anomaly

- The *New RR-Interval* contains a certain amount of noise but the signal quality is still fulfil the requirement of the cardiologist

- The *New RR-Interval* contains a certain amount of noise and the signal quality falls below the minimum acceptable level of the cardiologist. In which case, the *New RR-Interval* contains *Noisy Signal* anomaly

Among these five scenarios, this section is going to omit the *New RR-Interval* is free from anomaly scenario as the Figure 3.2, Figure 3.3, and Figure 3.4 in the Section 3.3.1 have already shown this scenario. In the rest of this section, it will mainly assess the *AID* with the other four scenarios. Figure 3.9 illustrates the *Correct RR-Interval* that is set and used as the ground truth for this scenario-based assessment. For the *Correct RR-Interval* and the rest signals used in this section are sampled with the sampling frequency of $360Hz$. The T_c is set to 0.8 as it is the preferred threshold from the cardiologist involve in this research.
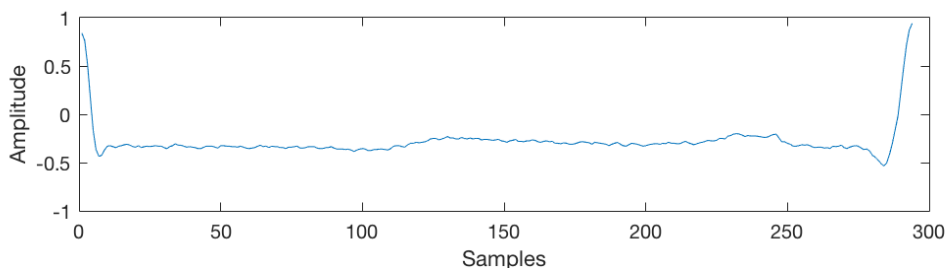
Figure 3.9: The *Correct RR-Interval* is set to be used in the scenario-based assessment

Scenario 1: When *No Signal* anomaly is presented

The scenario of detecting *No Signal* anomaly is presented is straightforward. It does not involve the RR-Interval prediction or signal classification. The *AID* con-
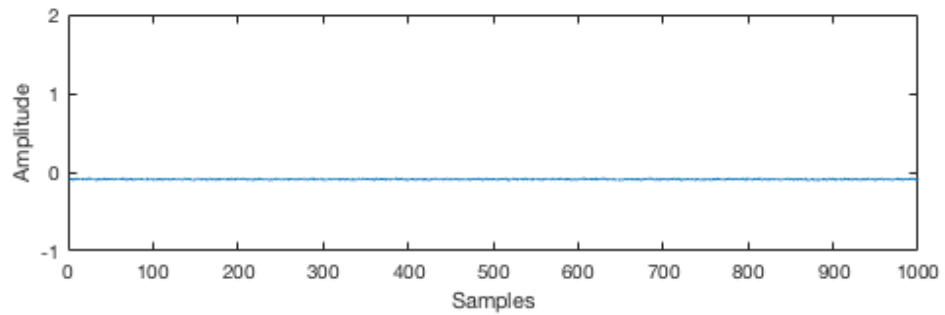
Figure 3.10: Scenario of *Blank Signal* anomaly is presented
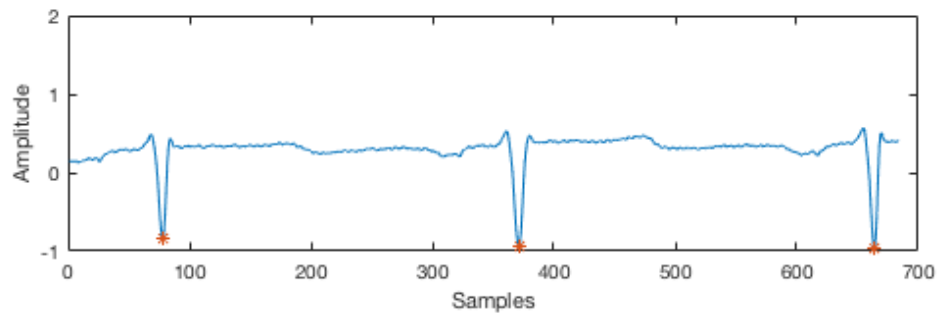
Figure 3.11: Scenario of *Inverted Signal* anomaly is presented

tinuously checks whether a sample is the R-Peak and counts the number of samples between two R-Peaks, which is subsequently compared to the maximum RR-Interval length, denoted as RR_{MAX}, set by cardiologist. If the number of samples between two R-Peaks exceeds RR_{MAX}, *AID* declares *No Signal* anomaly detected. Figure 3.10 illustrates a *No Signal* anomaly. It is assumed that the RR_{MAX} is set to 30 bpm, which means the number of samples between two R-Peaks is 720. In Figure 3.10, there are 1000 samples in total. However, none of the samples fulfils the R-Peak requirement so no R-Peak is detected for 1000 samples. As a result, *AID* declares this signal as with *No Signal* anomaly.

Scenario 2: When *Inverted Signal* anomaly is presented

Figure 3.11 illustrates a scenario of the captured signal containing the *Inverted Signal* anomaly. The *AID* starts with detecting the R-Peak within the signal. The detected R-Peaks are labelled by '*' in the figure. Then, *AID* extracts one RR-Interval as the *New RR-Interval* and aligns the *New RR-Interval* using the first R-Peak (as shown in Figure 3.12). The *AID* will then predicts the *Predicted RR-*
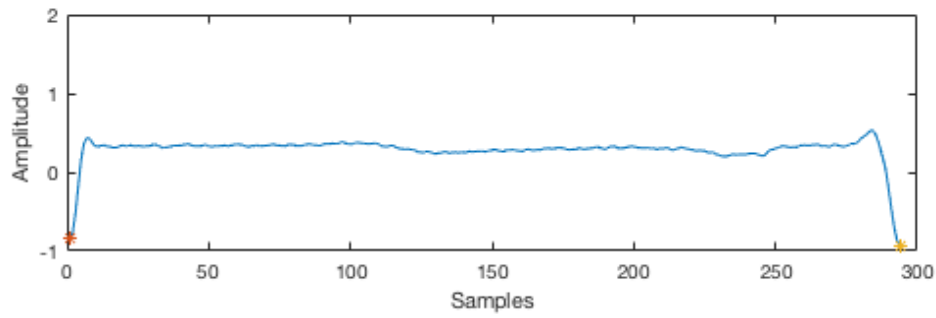
Figure 3.12: The *New RR-Interval* with *Inverted Signal* anomaly presented
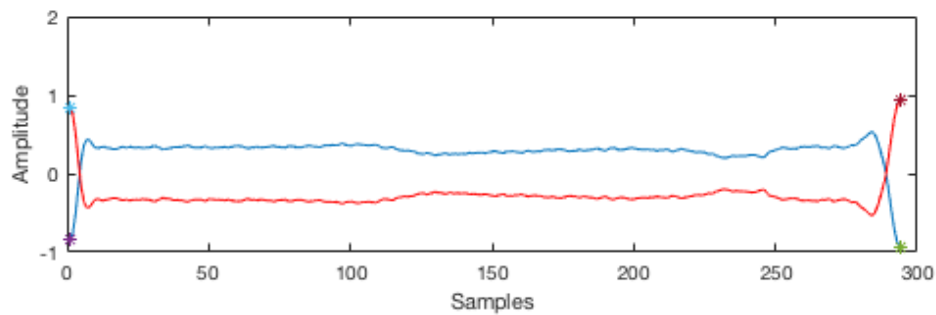
Figure 3.13: *AID* compares the *New RR-Interval* (blue) with the *Predicted RR-Interval* (red)

Interval using the *Correct RR-Interval* based on the length information of *New RR-Interval*. The *Predicted RR-Interval* is in the same size as the *New RR-Interval* (as shown in Figure3.13) and these two signals are compared. As the Pearson's correlation coefficient is -0.98, which is smaller than $-T_C$, the *New RR-Interval* is declared as containing the *Inverted Signal* anomaly.

Scenario 3: Noisy but still valid signal is presented

Figure 3.14 illustrate a *New RR-Interval* whose quality is still acceptable by cardiologist despite it contains noise. By comparing the *New RR-Interval* in Figure 3.14 and *Correct RR-Interval* in Figure 3.9, it returns the coefficient of 0.82. Such coefficient is higher than T_c. Subsequently, the *New RR-Interval* is classified as correct.

Scenario 4: When *Noisy Signal* anomaly is presented

Figure 3.15 illustrate a *New RR-Interval* whose quality is below the acceptance of a cardiologist. By comparing the *New RR-Interval* in Figure 3.14 and *Correct*
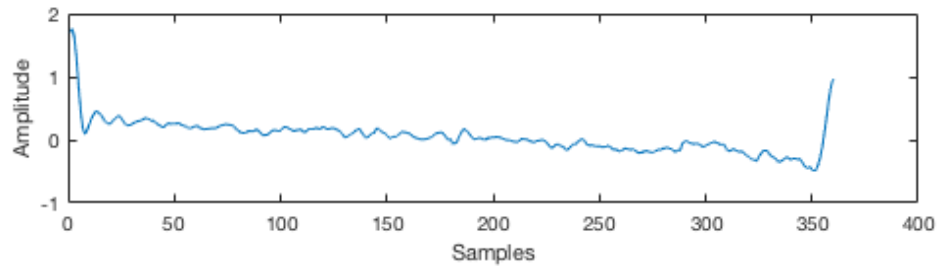
Figure 3.14: The *New RR-Interval* contains noise but the signal quality is acceptable by cardiologist
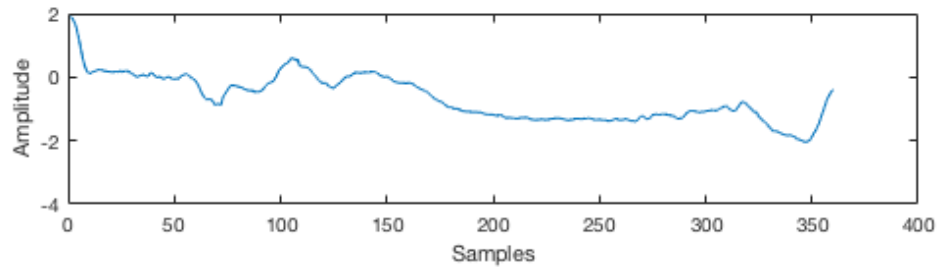
Figure 3.15: The *New RR-Interval* contains *Noisy Signal* and the signal quality falls below the acceptance of a cardiologist

RR-Interval in Figure 3.9, it returns the coefficient of 0.607 which is below T_c. Subsequently, the *New RR-Interval* is classified as containing *Noisy Signal* anomaly.

3.3.7 On-Mote Implementation

There are some modifications that are required to implement *AID* on a mote. These modifications are mainly due to the constrained on-mote environment. More specific, the calculation of Pearson's correlation coefficient involves the several operations of floating-point number which is not available on a mote. However, such limitation can be easily overcome by modifying the calculations. In this section, the modification to the calculation of Pearson's correlation coefficient will be introduced and the consequence of the modification will be discussed. The objective of such modification is to avoid the operations that require the floating-point number while minimise the performance degradation caused by the modification. Other than the calculation of Pearson's correlation coefficient, the *AID* can be implemented according to the pseudocode shown in Algorithm 1 without any modification. Therefore, they will not be discussed in this section.

In order to overcome the requirement of floating point number operation, the first step is to analyse which part of the Pearson's correlation coefficient is going to involve the floating-point number. Equation 3.2 below is the calculation of the Pearson's correlation coefficient, which is directly copied from Section 3.3.1. For the annotation of each variable, please refer to Equation 3.1 in Section 3.3.1.

$$r = \frac{\sum_{i=1}^{n}((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.2}$$

From Equation 3.2, the first operation that will involve the floating-point is the square-root calculation in the denominator. The result of the square-root can either be an integer or a floating-point number. To overcome this, a square operation is applied to both side of the equation so that the square-root is avoided but the equation still hold. However, the square operation will make any negative value become positive. Subsequently, the *AID* may not be able to detect *Inverted Signal* anomaly. To overcome this issue, the on mote implementation will first calculate the value of the numerator in Equation 3.2, and record sign of the numerator. Such recorded sign is used to decide the sign of the final result. As shown in Equation 3.2, the calculated value of the denominator will always be positive due to the square root operation. Thus, the sign of the final result of Equation 3.2 is decided by the numerator.

The modified equation is shown in Equation 3.3.

$$r^2 = \frac{(\sum_{i=1}^{n}((x_i - \bar{x})(y_i - \bar{y})))^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.3}$$

The next operation that involves the floating-point number is the fraction. The Pearson's correlation coefficient returns an index between -1 and 1, which will be a floating-point number. To overcome this, the approach is to scale the index up and only keep the integer part of the scaled number. For example, it is assumed that the index is 0.789012. To avoid the floating-point number, the index is scaled up by 100 times, which results as 78.9012. Then, only 78 is kept and 0.9012 is ignored. By

doing this, the floating-point number can be avoided. The advantage of such approach is that it is easy to implemented and does not require much more resource to run compared to original equation. On the other hand, the disadvantage is that there will be a certain level of accuracy degradation as only the integer part is kept. As the accuracy degradation is controlled by the scale factor, it is important to find a balance between overly scale up so that more resource is required, and insufficiently scale up so that accuracy suffered. The chosen scale factor in this research is 100 as it is believed that two digits after the decimal point is sufficient to classify the result. The final modified equation to the Pearson's correlation coefficient is shown in Equation 3.4.

$$r^2 * 100 = \frac{(\sum_{i=1}^{n}((x_i - \bar{x})(y_i - \bar{y})))^2 * 100}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.4}$$

Through the modification, the calculation of Pearson's correlation coefficient does not require the ability to perform floating-point number operation. However, due to the modification, some subsequent modification will need to be made to *AID*. It can be seen from Equation 3.4 that the index provided by Pearson's correlation coefficient is now changed from r to $r^2 * 100$. That means the classification threshold in *AID* will need to be changed accordingly. For example, the cardiologist, whom the author of this thesis consulted to, set the threshold T_C to 0.8 for the original *AID*. When *AID* is implemented on a mote, the threshold should now be $0.8^2 * 100$, which is 64.

3.4   Evaluation 1: Matlab Simulation

In this section, the evaluation of the proposed *AID*, which is carried out in Matlab, will be presented. The benefit of evaluating in Matlab is that a large scale of data from different subjects can be used to evaluate the detection accuracy. By using the large scale of data from different subject, the performance of the proposed *AID* can be fully evaluated and better demonstrated. The rest of this section will start with introducing the metrics used in this evaluation. How the evaluation is carried out is followed. The result will be presented at last.

3.4.1 Evaluation Metrics

To evaluate the detection accuracy, true-positive rate (abbreviated as TP), true-negative rate (abbreviated as TN), false-positive rate (abbreviated as FP), and false-negative rate (abbreviated as FN) are used as the evaluation metrics. Each of these metrics are explained below:

- **True-Positive Rate:** A true-positive result is defined as the the algorithm returns a positive result (containing usage-related anomaly) and such result is true. For example, *AID* reports that the Nth RR-Interval contains the usage-related anomaly. Then, by checking the signal annotation, it is confirmed that the Nth RR-Interval is anomalous. Based on this, the true-positive rate is defined as the number of true positive results divided by the total number of anomalies existed in the test signals (as shown in Equation 3.5).

$$TP = \frac{\sum Number\ of\ True\ Positive\ Results}{\sum Number\ of\ anomalies\ that\ existed} \qquad (3.5)$$

- **True-Negative Rate:** A true-negative result is defined as the the algorithm returns a negative result (free from usage-related) and such result is true. For example, *AID* reports that for the Nth RR-Interval, it is free from the usage-related anomaly. Then, by checking the signal annotation, it is confirmed that the Nth RR-Interval is anomaly free. Based on this, the true-negative rate is defined as the number of true-negative results divided by the total number of anomaly free RR-Intervals existed in the test signals (as shown in Equation 3.6).

$$TN = \frac{\sum Number\ of\ True\ Negative\ Results}{\sum Number\ of\ anomaly\ free\ RR\text{-}Interval\ that\ existed} \qquad (3.6)$$

- **False-Positive Rate:** A false-positive result is defined as the the algorithm returns a positive result (containing usage-related anomaly) but such result is a false result. For example, *AID* reports that the Nth RR-Interval contains the usage anomaly. However, by manually checking, the Nth RR-Interval is

free from anomaly. Based on this, the false-positive rate is defined as the number of false-positive results divided by the total number of RR-Intervals (as shown in Equation 3.7).

$$FP = \frac{\sum Number\ of\ False\ Positive\ Results}{\sum Number\ of\ RR\text{-}Interval} \qquad (3.7)$$

- **False-Negative Rate:** A false-negative result is defined as the the algorithm returns a negative result (free from usage-related anomaly) and such result is a false result. For example, *AID* reports that for the Nth RR-Interval, it is free from the usage anomaly. Then, by manually checking, the Nth RR-Interval contains usage anomaly. Based on this, the false-negative rate is defined as the number of false-negative results divided by the total number of RR-Intervals (as shown in Equation 3.8).

$$FN = \frac{\sum Number\ of\ False\ Negative\ Results}{\sum Number\ of\ RR\text{-}Interval} \qquad (3.8)$$

In addition to the evaluation metrics for detection accuracy, recognition rate (abbreviated as Rec) is used as the evaluation metric for the anomaly recognition accuracy. The recognition rate is defined as the number of correctly recognised anomalies divided total number of true positive detections (as shown in Equation 3.9). For example, it is assumed that there are 100 *Noisy Signal* anomalies existed in the test signal and the TP is 100%. That means all existed anomalies have been correctly detected. Among these detected anomalies, if the *AID* report 90 out of 100 anomalies are *Noisy Signal* anomaly, the recognition rate is 90%.

$$Rec = \frac{\sum Number\ of\ correctly\ identified\ anomalies}{\sum Number\ of\ True\ Positive\ Results} \qquad (3.9)$$

3.4.2   Evaluation Setup

This evaluation aims at showing the detection accuracy achieved by *AID* against each type of usage-related anomalies, the ability to achieve similar detection ac-

curacy across different patients, and how these results is compared with existing ECG anomaly detection approach. To achieve these objectives, the following procedures are followed in this evaluation. In general, test signals with known ground truth for the detection results will be fed to the algorithms. The detail of how the test signals are generated and how the ground truth is obtained will be introduced in Section 3.4.3. As the ground truth for the detection result is known, the actual detection result will be compared with the known ground truth so that the evaluation metrics can be calculated. More specifically, this evaluation is divided into three experiments. For each experiment, only one type of usage-related anomaly will be presented in the test signals. Thus, it is able to evaluate the detection accuracy against each usage-related anomaly. Meanwhile, the anomalous signals and the anomaly-free signals are mixed together and fed to the algorithm. Therefore, there is no need to have a separate experiment to evaluate the false positive rate. To evaluate the ability to achieve similar detection accuracy across different patients, each experiment will feed 20 test signals from different patients to the algorithm individually with the evaluation metrics being calculated and stored for each individual test signals as well as the overall achieved results. Thus, the outcome from each experiment will not only give the evaluation metrics achieved by the algorithm overall, but also give the metrics achieved with each individual test signal. Then, the results achieved with each test signal will be compared with each other. The detail of how the results are analysed will be covered in Section 3.4.4. Finally, in order to compare the results achieved by the *AID* to existing approach, the *AWDD* is chosen to be compared with. Though the review of existing works in Section 2.2.3, it has been identified that the *AWDD* are the approach that is designed for the signals from ECG-Based HC-BSNs, and it may work for detecting the usage-related anomaly. Thus, *AWDD* is chosen for comparison in this evaluation. To perform the comparison, the signals, which are fed to the *AID* during each experiment, will also be fed to the *AWDD* at the same time. The detection results reported by the *AWDD* will also be compared with the ground truth so that the evaluation metrics can be calculated. Exact the same procedure will also be followed to evaluate the result variation when the signals from different patients are fed to the algorithm.

Parameters	Values
RR_C	First RR-Interval from each test signal
L_C	Measured from the RR_C
T_c	0.8 (set according to cardiologist)
T_R	Measured from the RR_C plus 10% tolerance
RR_{MAX}	720 (equal to 30 bpm when sampling frequency is $360Hz$)

Table 3.1: Parameters used to evaluate the *AID*

Overall, by following the above procedure, it is able to achieve the objectives in this evaluation.

In the Algorithm 1, there are five parameters that need to be pre-determined before the detection. To carry out the experiments, these parameters used to configure the *AID* are shown in Table 3.1. For the **Correct RR-Interval** (RR_C), we use the first RR-Interval from each test signal as they are free from usage-related anomaly. Then, the length of the **Correct RR-Interval** (L_C) is set by measuring the length of the defined RR_C. For the T_c, we use the value of 0.8, which is recommended by the cardiologist based on her signal quality standard. For the R-Peak detection threshold (T_R), we measure the difference between the amplitude of the R-Peak and the amplitude of its next data sample. Then, we give it 10% tolerance. Such tolerance is given considering the possible variation of each RR-Interval. For example, if the measured amplitude difference is 50, then the T_R is set to 45 as there is 10% tolerance. Finally, the allowed maximum RR-Interval length (RR_{MAX}) is set to 720, which also followed the recommendation from the cardiologist.

3.4.3 Procedure To Generate Test Signals

In order to obtain the evaluation metrics accurately and objectively, it requires the evaluation environment being controlled. The controlled evaluation environment means the exact location of where the anomaly occurs, and the type of the anomaly in the evaluation signals are known before start the evaluation. However, as reviewed in Section 2.2.4, existing databases are not suitable for the purpose of this research due to the signals being annotated for heart condition instead of usage-related anomaly. Thus, alternative solution need to be found. For the research in ECG signal related domain, the most widely acknowledged and used approach

to generate test signal is proposed by Moody et al. [58]. In a short summary to Moody's approach, they began with recording different type of noise signals which may appear in ECG signals in reality. Then, they inject the recorded noise signal into the source signals, which are from MIT-BIH Arrhythemia Database [57], by adding the noise signal with the source signal in a sample-to-sample basis. Before the injection, the strength of the noise is controlled by the researchers. Thus, it is able to provide a controlled evaluation environment. Moody et al. also claimed that their approach can generate the test signals with sufficiently realistic to represent the real world scenario. However, in this research, it is not expected that the usage-related anomaly is appeared for the entire duration of the source signal. Instead, it is expected that the anomalous signal and anomaly-free signal are mixed together. Thus, Moody's approach can not be directly adopted in this research.

Inspired by Moody's approach, the test signals used to evaluate *AID* are also generated by injected usage-related anomalies into the source signals which is free from usage-related anomaly. As it is desired that the usage-related anomalies are mixed with anomaly-free signals for each test signal, the test signal generator will randomly select the point where the usage-related anomaly is injected, and the duration which the injected anomaly will last for. At the same time, the point where the anomaly is injected and the duration which the anomaly will last for will be recorded and used as the ground truth to decide whether the detection result is a true result or a false result. With this two parameter, the usage-related anomaly will be injected into the source signal by adding the anomaly and the signal together. However, as the type of anomaly may affect the result of R-Peak detection, for a given source signal, the actual number of detections that will be run may vary depending on which type of usage-related anomaly is injected into the source signal. Furthermore, as the usage-related anomaly is injected into the source signal at a random point for a random duration, the number of anomalies in the test signals may also vary between each experiment. For this test signal generation approach, it is important to select the appropriate source signal database and the anomaly database.

Page 68

In this research, the MIT-BIH Arrhythmia Database [57] is used as the source signal database as it is widely used by the researcher in ECG related researches. The MIT-BIH Arrhythmia Database was constructed during the 1980s after Moody et al. recognised that a more suitable set of well-characterized ECG signals was needed for research groups in both academia and industry to carry out reproducible and objectively comparable evaluation for the arrhythmia detectors. Since the 1980, the MIT-BIH Arrhythemia Database has been used at about 500 sites worldwide [57]. The database contains 48 half-hour excerpts of ECG recordings obtained from 47 subjects who were studied by the BIH Arrhythmia Laboratory between 1975 and 1979 [57]. Those subjects included 25 male and 22 female whose age were between 23 to 89 years old. The ECG recordings were digitised using 11-bit Analogue-to-Digital Converter (ADCs) with $360Hz$ sampling frequency. For this database to be used in this research, 20 out of 48 ECG recordings are manually selected depending on the signal quality (e.g low noise). As each cardiologist or researcher may have his/her own requirement to the signal quality, not every signal in the database has met the quality requirement set by the cardiologist whom we consulted to during this research. Thus, these 20 signals are selected and meet the quality requirement set by our cardiologist.

For the anomaly database, the following methods are used to obtain each type of usage anomaly signals. For the *No Signal* anomaly, it is captured by Shimmer2r [2] without attaching any electrode. The data capture parameters in Shimmer2r (e.g sampling frequency) are set exactly the same as in the MIT-BIH Arrhythmis Database. *Inverted Signal* anomaly is obtained by inverting part of the signal from MIT-BIH database. For *Noisy Signal* anomaly, noise signals from MIT-BIH Noise Stress Test database [58] are used. In the database, it contains three common types of noise that can be seen in an ECG recordings: baseline wander, electrode motion artifact, and muscle noise. Although Moody et al. acknowledged that the power line interference is the forth source of noise, they ignored the power line interference due to being easily removed [58]. As the noise strength can play an important role in the result, noisy signals with SNR of $-10dB$, $-6dB$, and $0dB$,

which are believed to cover the possible noise strength range in reality according to our cardiologist, are injected to the source signals.

3.4.4 Analysis Method

In order to achieve the evaluation objectives, the results obtained from each experiment will need to be analysed. In this evaluation, the basic result presenting and comparison method is to put the obtained evaluation metrics into a table, which is an intuitive way to demonstrate the results achieved by different algorithms. On top of the basic result presenting method, the following two methods will be used to demonstrate the result variation between test signals from different patients, and the significance of the obtained results.

Box-Whiskers Plot

Box-Whiskers Plot (abbreviated as Boxplot in the rest of this thesis) is a widely used non-parametric way to show the data distribution graphically. For the Boxplot, the horizontal axis usually represents the items plotted in the graph whereas the vertical axis represents the values where the data is distributed. The plot for each item will show its interquartile range using a box with the median value in the data shown by a red line in the box. The bottom of the box represents the first quartile value (median of the lower half of the data set) and the top of the box represents the third quartile value (median value of the upper half of the data set). A whisker, which is drawn through the box vertically, shows the overall data range with the top of the whisker being the maximum value and bottom of the whisker being the minimum value. Except for the box and whisker, there may be some other symbols (in this research, it is red '+' symbol) that are drawn above the maximum value or below the minimum value. These symbols represent the outliers in the data set. The outlier is usually defined as one and half time of the interquartile range above the third quartile value or below the first quartile value. In this research, this definition of the outlier is adopted when plotting the Boxplot. When comparing the results achieved by two algorithms, the outlier will not be considered as they are deviated from the majority of results. Instead, the outliers will be analysed

individually to check what causes them to deviate from the majority of results so that a better understand of the algorithm's behaviour can be studied.

Wilcoxon Rank-Sum Test

Data obtained from multiple measurements may be subject to error [48]. Thus, it is necessary to apply statistical significance test to determine whether the observed difference in the results achieved by two algorithms is due to random errors. For example, the overall detection accuracy achieved by *AID* may be higher than *AWDD*. However, such difference can be caused by the detection accuracies across different test signals achieved by two algorithms are too close to each other so that the difference is purely random. In order to assess the significance of the difference achieved by two algorithms, Wilcoxon Rank-Sum Test (also known as U-Test) [87] will be used in this evaluation. The Wilcoxon Rank-Sum Test is a non-parametric test. The advantage of Wilcoxon Rank-Sum Test is that it does not require the assumption of the test data being normally distributed whereas other statistical tests (e.g T-Test) require such assumption. Due to this advantage, it can avoid the verification of whether the test data is normally distributed. Thus, Wilcoxon Rank-Sum Test is selected to perform the statistical significance test in this research. The outcome from the Wilcoxon Rank-Sum Test will show the p-Value. To assess whether the difference is significant, a widely used and acknowledged threshold of 0.05 for the p-Value is used in this research. If the p-Value is below or equal to 0.05, the difference is considered to be significant. Otherwise, the difference is considered to be not significant.

3.4.5 Evaluation Result

Experiment 1: *No Signal* Anomaly Injected

In Experiment 1, only the *No Signal* anomaly is injected into the source signals. The overall results achieved by *AID* and *AWDD* are shown in Table 3.2. Overall, there are 4047 detections have been performed with 200 injected anomalies. When *No Signal* anomaly is injected, *AID* can improve the TP and TN rate by 66.67% and 5% respectively compared to *AWDD*. *AID* also reduces the FP and FN rate by

	TP	TN	FP	FN	Rec
AWDD	33.33%	93.01%	6.81%	0.39%	-
AID	100%	98.01%	0.52%	0%	100%
Improvement	66.67%	5%	6.29%	0.39%	-
p-Value	6.9e-09	0.004	0.003	1.1e-06	-

Table 3.2: Overall result achieved for *AWDD* and *AID* when only the *Blank Signal* anomaly is injected

6.29% and 0.39% respectively. It can correctly recognise all injected *No Signal* anomalies. The calculated p-Values for each rate are all well below 0.05. It indicates that the improvement achieved by *AID* over *AWDD* is statistically significant.

Figure 3.16 shows the boxplot of TP, TN, FP, FN when the *Blank Signal* anomaly is injected into the source signals. It can be seen from the plots that the *AID* can achieve similar results for the true-positive and false-negative rates across 20 subjects. That means for the *Blank Signal* anomaly, *AID* can detect all possible *Blank Signal* anomaly without any omission. For the true-negative and false-positive rate, if not considering the statistical outlier, the variation between difference subjects are also less than *AWDD*.

Experiment 2: *Inverted Signal* Anomaly Injected

In Experiment 2, only the *Inverted Signal* anomaly is injected into the source signals. The overall results achieved by *AID* and *AWDD* are shown in Table 3.3. In total, 4019 detections have been performed with 271 *Inverted Signal* anomalies being injected. Overall, *AID* has managed 80.07% TP, 12.86% TN, 6.39% FP, and 7.49% FN improvement over *AWDD*. Also, it managed to recognise 96.31% of all the injected anomalies. By carrying out the Wilcoxon Rank Sum Test, the p-Values of each rates are all below 0.05. Subsequently, it can be said that the detection accuracy improvements over *AWDD* are significant.

The boxplot for the results in Experiment 2 is shown in Figure 3.17. Except the statistical outliers, the *AID* has achieved better TP and FN with less variation between different test subjects than the *AWDD*. For those non-outliers TP and FN, the *AID* can achieve nearly 100% and nearly 0% respectively. On the other hand,

Figure 3.16: Boxplot of TP, TN, FP, FN showing the maximum, minimum, median upper quartile, lower quartile, and interquartile range with the statistical outliers (plus sign) when *Blank Signal* anomaly is injected to the source signals

	TP	TN	FP	FN	Rec
AWDD	18.45%	84.73%	7.02%	7.75%	-
AID	98.52%	97.59%	0.66%	0.26%	96.31%
Improvement	80.07%	12.86%	6.39%	7.49%	-
p-Value	$2.3e\text{-}08$	$2.5e\text{-}04$	0.002	$2.8e\text{-}06$	-

Table 3.3: Overall result achieved for *AWDD* and *AID* when only the *Inverted Signal* anomaly is injected

the TP and FN achieved by *AWDD* have much wider variation across different test subjects. For the TN and FP, the *AWDD* achieved slightly less variation across different test subjects than the *AID*. However, the result from *AID* is generally better than the *AWDD* despite the slightly wider range.. For example, the minimum value from *AID* is almost equal to the upper quartile achieved by *AWDD*.

Experiment 3: *Noisy Signal* Anomaly Injected

In Experiment 3, the *Noisy Signal* anomaly is injected into the source signals. The noise strength is varied between $-10dB$ and $0dB$. The overall results achieved by *AID* and *AWDD* are shown in Table 3.4. In general, the results show that when the noise is much stronger than the signal, *AID* can achieve much better detection accuracy than *AWDD*. When the noise strength decreased, the TP rate decreases an FN rate increases. More specific results are presented below for each noise strength level.

When the noise strength is at -10 dB, the *AID* can improve the detection accuracy by 56.03%, 3.78%, 3.67%, and 1.55% respectively for TP, TN, FP, and FN. The Wilcoxon Rank Sum Test returns a smaller than 0.05 p-Value for each rate. That means the improvements achieved by the *AID* are all statistically significant. From the boxplot of these results shown in Figure 3.18, it can be seen that the results from *AID* are better than *AWDD*.

When the noise strength is -6 dB, the *AID* can also manage an improvement of 52.3%, 4.08%, 4.07%, and 1.42% respectively for the TP, TN, FP, and FN. The p-Values from Wilcoxon Rank Sum Test also show that the improvement for each rate is significant. The boxplot of the results is shown in Figure 3.19. Similar observation

Figure 3.17: Boxplot of TP, TN, FP, FN showing the maximum, minimum, median upper quartile, lower quartile, and interquartile range with the statistical outliers (plus sign) when *Inverted Signal* anomaly is injected to the source signals

Noise Strength		TP	TN	FP	FN	Rec
-10dB	AWDD	31.91%	91.18%	6.25%	2.32%	-
	AID	87.94%	94.96%	2.58%	0.77%	99.19%
	Improvement	56.03%	3.78%	3.67%	1.55%	-
	p-Value	5.9e-08	0.002	0.001	2.5e-06	-
-6dB	AWDD	32.33%	90.76%	6.66%	2.33%	-
	AID	84.21%	94.84%	2.59%	0.9%	99.11%
	Improvement	52.3%	4.08%	4.07%	1.42%	-
	p-Value	5.9e-08	0.002	0.002	1.1e-04	-
0dB	AWDD	29.81%	90.11%	7.42%	2.44%	-
	AID	51.92%	92.96%	2.64%	2.83%	100%
	Improvement	22.11%	2.85%	4.78%	-0.39%	-
	p-Value	5.3e-05	0.003	0.001	0.67	-

Table 3.4: Overall result achieved for *AWDD* and *AID* when only the *Noisy Signal* anomaly is injected with strength between $-10dB$ to $0dB$

to the results from the noise strength of -10 dB can be seen from Figure 3.19. The *AID* generally perform better than *AWDD* with less variation across different subjects.

When the noise strength reduced to 0 dB, the improvements start to decrease. The *AID* achieved 22.11%, 2.85%, and 4.78% respectively for TP, TN, and FP. However, the FN is decreased by -0.39% compared to *AWDD*. Through the Wilcoxon Rank Sum Test, the p-Values of $5.3249e-05$, 0.0029, and 0.0012 for TP, TN, and FP respectively show that the improvements for these three rates are significant. However, the p-Value for the FN is 0.6749, which means that the performance decrease compared to *AWDD* is not statistically significantly.

3.4.6 Analysis of Outlier Results

Through the evaluation, the detection accuracy for *Blank Signal*, *Inverted Signal*, and *Noisy Signal* with noise stronger than the signal (below 0 dB) is as what is expected by the author of this thesis. However, through the boxplots of the evaluation result, there are constantly some statistical outliers for *AID*. By inspecting each boxplot's data, it is observed that no matter which type of anomaly is injected, the outliers are always happened with the following source signals: MIT-BIH#106, MIT-BIH#214, MIT-BIH#217, and MIT-BIH#101 with the MIT-BIH#106 being the worst
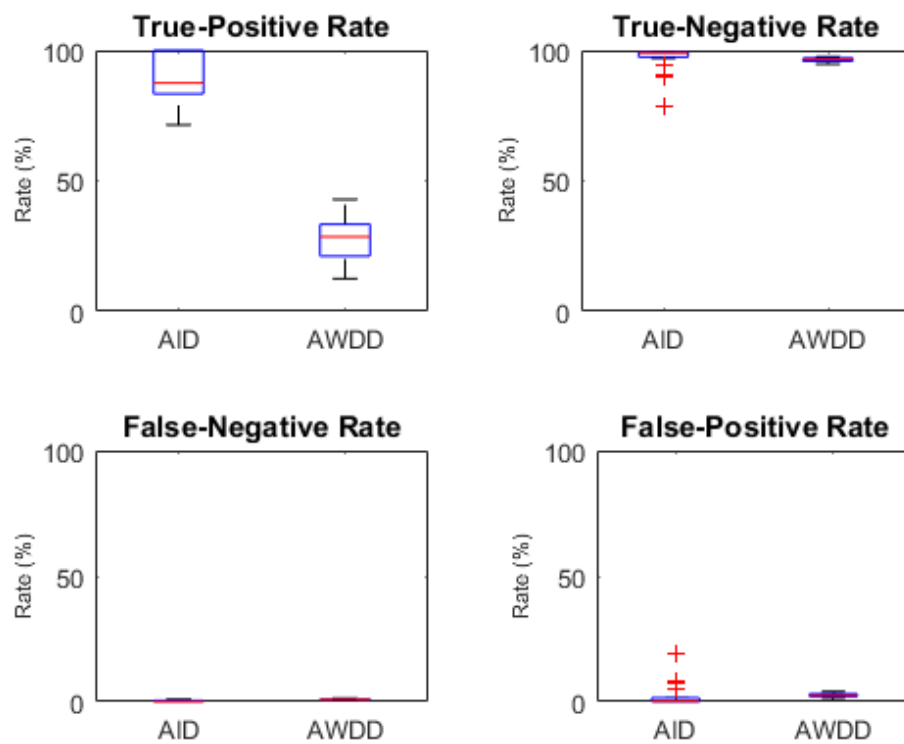
Figure 3.18: Boxplot of TP, TN, FP, FN showing the maximum, minimum, median upper quartile, lower quartile, and interquartile range with the statistical outliers (plus sign) when *Noisy Signal* anomaly with the noise strength of $-10dB$ is injected to the source signals
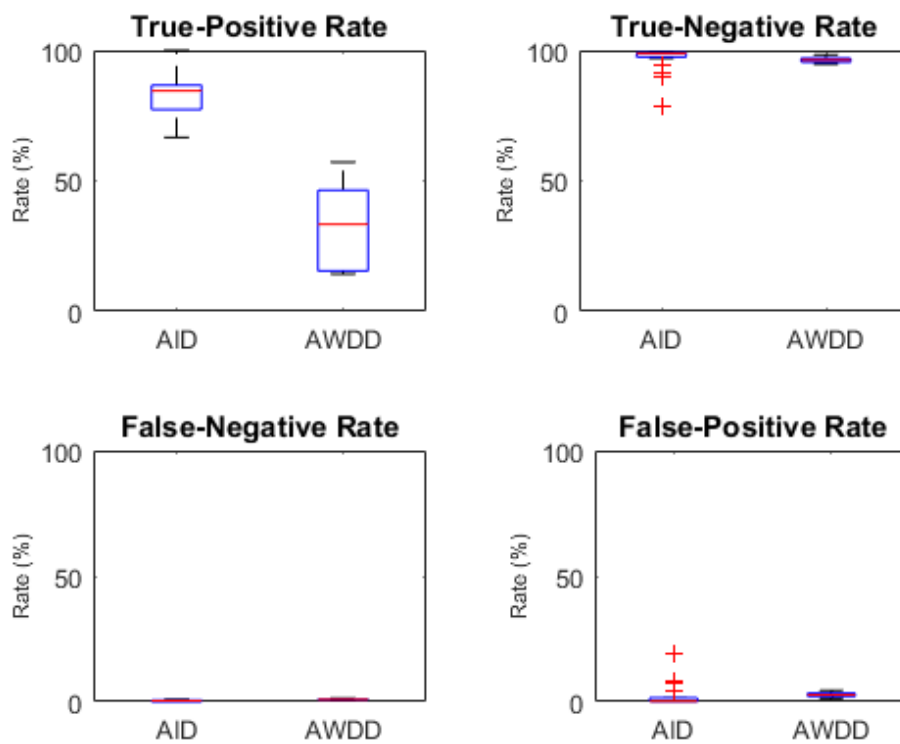
Figure 3.19: Boxplot of TP, TN, FP, FN showing the maximum, minimum, median upper quartile, lower quartile, and interquartile range with the statistical outliers (plus sign) when *Noisy Signal* anomaly with the noise strength of $-6dB$ is injected to the source signals
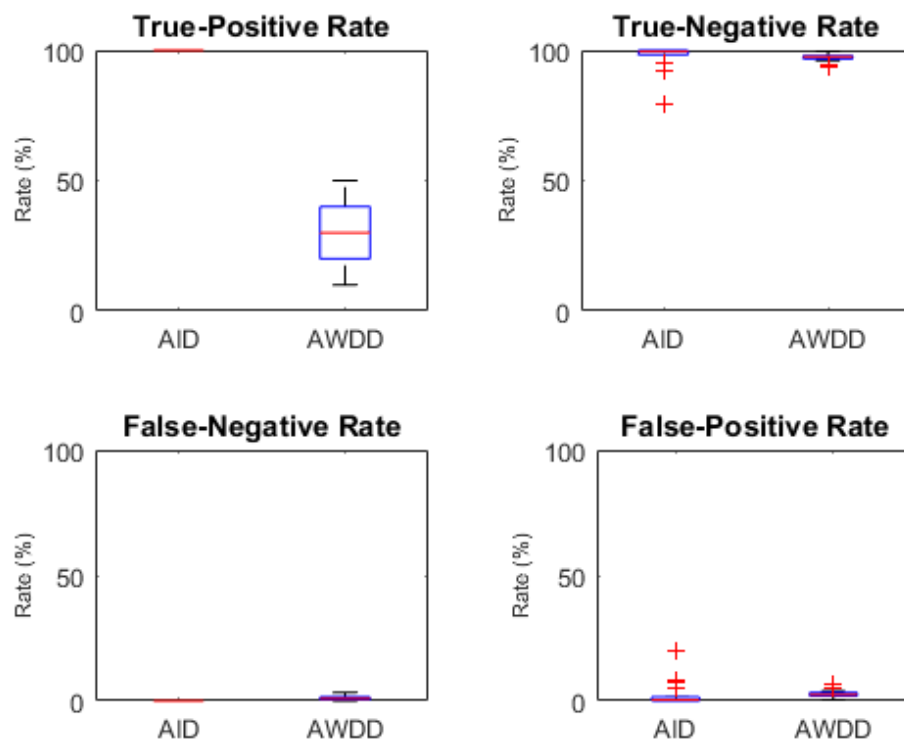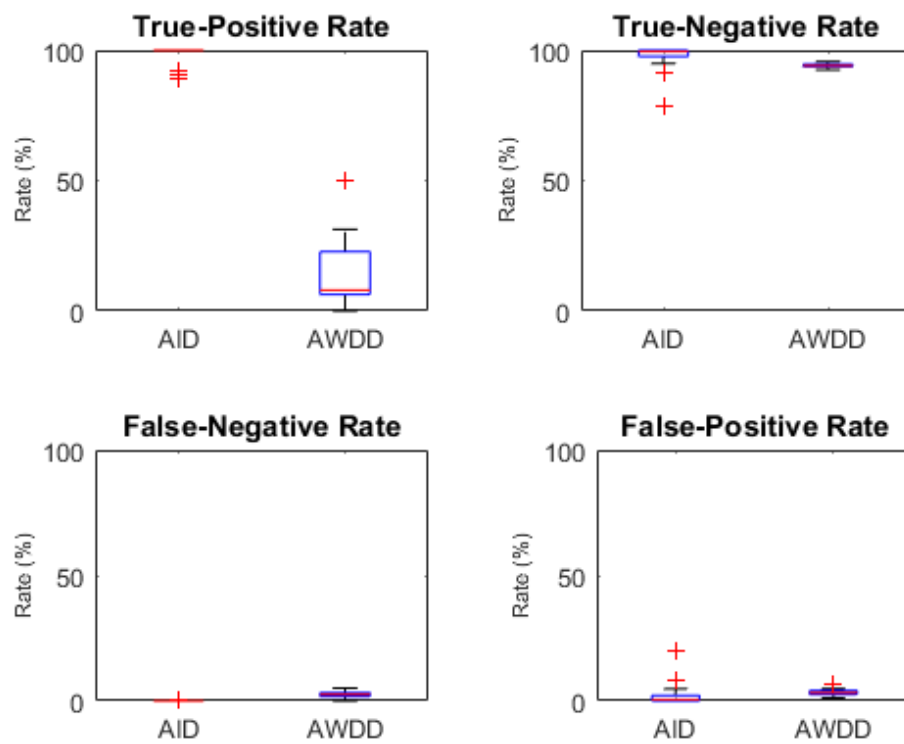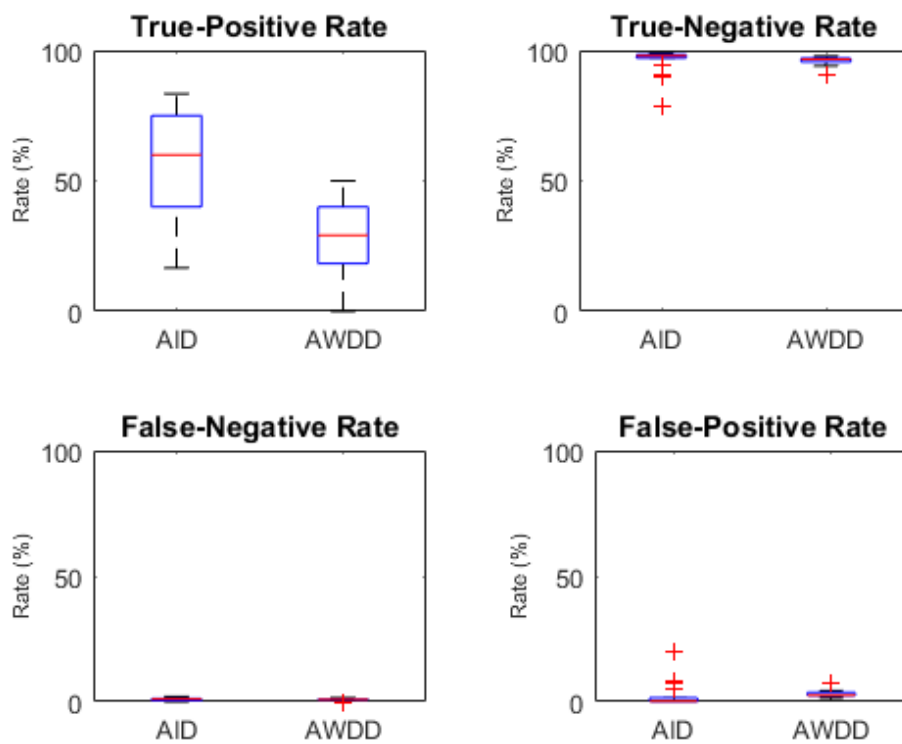
Figure 3.20: Boxplot of TP, TN, FP, FN showing the maximum, minimum, median upper quartile, lower quartile, and interquartile range with the statistical outliers (plus sign) when *Noisy Signal* anomaly with the noise strength of $0dB$ is injected to the source signals

outlier and MIT-BIH#101 the least worst outlier. To further investigate these cases, an extra analysis has been carried out to explore the causes of these outliers.

	Count	Total Detection Count
Heart Condition	18	2570
Inaccurate **Correct RR-Interval**	88	2570
Inaccurate R-Peak detection	9	2570
Noisy source signal	5	2570

Table 3.5: The causes of FP detection

For the FP result, the investigation result is categorised into 4 groups (shown in Table 3.5). The most cause of the outlier result is due to inaccurate **Correct RR-Interval**. In the implementation of the *AID*, a feature is added to update the preset **Correct RR-Interval** if the **New RR-Interval** has been classified as free from usage-related anomaly. Such feature aims at making the *AID* tolerate the possible change of the ECG signal for long-term. However, as the threshold setting allows some tolerance, the signal error may accumulate with each update. As a result, the stored correct signal may become incorrect. For example, if the correlation between a **New RR-Interval** and the **Predicted RR-Interval** is 0.81 which is higher than T_c, the **New RR-Interval** will become the **Correct RR-Interval**. However, if such sequence is continuing and each time the correlation is 0.81, the **Correct RR-Interval** will become significantly different to the originally set **Correct RR-Interval**. Thus, if using the distorted **Correct RR-Interval** to perform the detection, the detection result will become inaccurate. Therefore, high FP results are generated. By removing such **Correct RR-Interval** update feature, the result shows that it is able to remove those 88 FP result. Subsequently, the performance of AID can be further improved. However, the long-term side effect of removing such feature is not investigated so far due to lack of long-term signal. The second most cause of the FP results is the signals with heart conditions. The source signal database used in this evaluation is designed to evaluate the performance of arrhythmia detectors. Subsequently, all the signals in this database contain different type of heart conditions. When the *AID* is designed, it is desired to ignore the anomalous signals caused by heart conditions. However, some of the heart conditions can significantly distort

the signals so that the *AID* classifies them as anomalous. With the current design of the *AID*, such issue may not be able to addressed as the *AID* only uses one **Correct RR-Interval** to justify the correctness of the new signal due to resource consumption consideration.

3.5   Evaluation 2: On-Mote Evaluation

This section is going to present the on-mote evaluation. Like in previous section, this section will start with introduce the objectives of the on-mote evaluation followed by the metrics which are used in this evaluation. Then, how the evaluation is carried out will be introduced. Finally, the evaluation result will be presented.

3.5.1   Evaluation Objective

There are three main objectives for the on-mote evaluation. The first objective is to evaluate the claim of the *AID* being capable to perform the detection on a mote at run-time. The second objective is to evaluate the degradation caused by the on-mote implementation without using the floating-point number operation (introduced in Section 3.3.7). The third objective is to evaluate the overhead generated by the *AID*. It includes the computation overhead, programme ROM overhead, and the run-time RAM overhead.

3.5.2   Evaluation Metrics

On top of the evaluation metrics that are used in the Matlab simulation to evaluate the detection accuracy, the following metrics are used for evaluate the overhead:

- **Compiled Binary Size (unit: kByte):** To evaluate the overhead in terms of the firmware size, which is generated by AID, the compiled binary size is used. It is the size of the .ihex file which is compiled from the C-Language code of the firmware.

- **Consumed RAM Size (unit: kByte):** To evaluate the overhead in terms of the *AID*'s RAM consumption during run-time, the consumed RAM size is used. It is calculated through the running parameters of *AID*. The parameters

include the sampling resolution, sampling frequency, and estimated heart-rate. Equation 3.10 is used to calculate the consumed RAM size.

$$Overhead = (((60/HR * Fs) * 2 * Resolution)/8)/1024 \qquad (3.10)$$

where:

HR: Heart Rate (bpm)

Fs: Sampling Frequency (Hz)

Resolution: Quantisation Resolution (bits)

Overhead: RAM Overhead (kByte)

- **Time to Perform One Detection (unit: ms):** To evaluate the overhead in terms of computation, the time to perform one detection by *AID* is used. It is obtained by counting the time spent from the detection start till the detection finished.

3.5.3   Evaluation Setup

Based on the same reasons as the Matlab simulation, it is still desired that the evaluation environment for the on-mote evaluation is controlled. Therefore, the same fault injection approach as being used during the Matlab Simulation will be adopted. However, due to the different purpose of this evaluation, the procedure of generating the test signal is modified slightly in this evaluation. As this evaluation does not need to evaluate the detection accuracy between different patients, only signal MIT-BIH#100 from the MIT-BIH database will be used as the source signal. Also, as the aim of evaluating the detection accuracy is to check the degradation caused by on-mote implementation, instead of injecting each type of anomalies separately, this evaluation will inject all three types of usage-related anomalies together at a random point with random duration. To make the results from the mote and from the Matlab comparable, the same test signal will be fed to the algorithm

Page 82

implemented on-mote, and the algorithm implemented in Matlab with the former one being without floating-point operations whereas the later one is with floating-point operations.

To carry out the evaluation, *AID* is implemented on a widely used mote platform (both hardware and software). The hardware platform used in this evaluation is the T-Mote Sky [19] which uses exactly the same MCU as the Shimmer2r with the same operating frequency, and also has the same amount of on-board ROM and RAM. On top of sharing the same core hardware, T-Mote Sky is widely used by researchers. Therefore, the low level software support (e.g peripheral drivers, interface with computer, toolset on computer etc.) is much more mature than the Shimmer2r. The software platform used in this evaluation is the Contiki OS [20]. Contiki is one of the widely used operating system in sensor network area. With its conventional programming technique based on C-Language, it simplifies the implementation of this evaluation. Furthermore, the Contiki OS comes with a matured simulator called Cooja [62] which further simplify the debugging process of the implementation.

During the evaluation, $1,000$ seconds of signal from MIT-BIH#100 with mixed anomalies injected are segmented into five hundreds of 2-second signals and loaded onto the mote. The segmentation is due to the on mote memory limitation. It is not possible to load entire $1,000$ seconds of signal onto a mote. The evaluation is then carried out on the mote with the result collected by a computer through the on-board serial port. The collected results will then be compared with the ground truth so that the evaluation metrics for detection accuracy can be obtained. In addition to the detection accuracy, the size of the compiled firmwares, both with and without the *AID*, will be measured so that the ROM overhead can be evaluated. To evaluate the computational overhead, a counter is set to count the number of machine cycle for each detection. The counter starts counting when the *AID* starts the detection, and stops as soon as the detection is done. The number of machine cycle is then converted to time in millisecond.For the RAM overhead, it may vary depending on each specific detection as the *AID* will need to store differ-

ent length of signal due to the possible heart-rate variation. Here in this evaluation, the RAM overhead will be analysed theoretically.

3.5.4 Evaluation Result

The result of on-mote evaluation for detection accuracy is shown in Table 3.6. In total, 1264 detections have been performed on mote and 621 anomalies have been injected at random location. Overall, *AID* achieved 98.60% TN, 98.71% TP, 0.63% FN and 0.71% FP on a mote. On the other hand, with the same signal being fed to Matlab simulation, it returns 99.53% TN, 99.19% TP, 0.4% FN and 0.24% FP. For these results, there is not much differences in TP and TN achieved on-mote and in Matlab. Therefore, it can be concluded that the on-mote implementation does not have significant degradation in terms of the TP and TN. On the other hand, the differences in FP and FN are greater than the differences in TP and TN (the difference is about 50% for FN and 300% for FP). However, if we look at the raw values used to calculate these rates, the on-mote implementation only reports 6 more FP detections and 3 more FN detections compared to the Matlab simulation. The difference in raw value is not much different compared to those rates. Thus, it is also believed that the on-mote implementation does not cause much degradation.

	Result from Matlab	Result from on-mote
True-Positive Rate	99.19%	98.71%
True-Negative Rate	99.53%	98.60%
False-Positive Rate	0.24%	0.71%
False-Negative Rate	0.4%	0.63%

Table 3.6: Detection accuracy comparison between the Matlab-based evaluation and the Mote-based evaluation

The result of on-mote evaluation for overheads is shown in Table 3.7. When compiling the firmware without the code of *AID*, the .ihex file size is 61.7kB. When integrating *AID* to the firmware, the compiled .ihex file increases to 64kB. That means it generates 2.3kB (3.72%) of overhead in terms of code size. During the run-time,

depending on the heart rate (HR) and sampling frequency (Fs), the storage of new captured signal and correct signal may vary. Assuming the lowest acceptable heart rate (HR) is 30 beats per minute (bpm) and the signal is sampled at 256Hz, the RAM consumption of the algorithm will be around 2 kByte (20% of total RAM on Shimmer2r or T-Mote Sky). On the other hand, if it is assumed that the highest acceptable HR is 120 bpm, the RAM consumption of the algorithm will be around 0.5 kByte (5% of total RAM on Shimmer2r or T-Mote Sky). It's also measured that *AID* requires around 80 ms to perform one detection to a RR-Interval whose typical duration is around 1 second (heart rate around 60 bpm). That means if the acceptable HR is vary between 30 bpm to 120 bpm, the computational overhead will be varied between 4% to 16% of the length of one RR-Interval.

Type of overhead	Generated Overhead	Overhead Percentage
Firmware Overhead	2.3 kBytes	3.72% of total ROM
RAM Overhead (Max; HR=30bpm)	2 kBytes	20% of total RAM
RAM Overhead (Min; HR=120bpm)	0.5 kBytes	5% of total RAM
Computational overhead	80 ms	8% if HR is 60 bpm

Table 3.7: The evaluation result for the overheads generated by *AID*

3.5.5 Discussion

At the beginning of this section, three evaluation objectives have been proposed. In order to show the achievement, each of the objectives will be reviewed here using the evaluation result as evidence.

The first objective is to evaluate the claim of the *AID* being capable to perform the detection on a mote at run-time. During the evaluation, the *AID* has been implemented based on Contiki operating system, and compiled to both Shimmer2r BSN Mote and T-Mote Sky. The calculated code size of 2.3 kBytes is well within the available programme ROM size. The RAM consumption of the *AID* can vary between 0.5 kBytes and 2 kBytes depending on the heart-rate. Such RAM consumption is also within the available RAM on a mote. Furthermore, the modification made to

the *AID* has removed the requirement to floating-point number operation. Thus, based on these evidences, it can be said that the proposed *AID* is capable to run on a mote. The evaluation of the computational overhead has shown that the *AID* can perform one detection against one RR-Interval for around 80 ms, which is well below the possible shortest RR-Interval length. Thus, it can also be said that the *AID* is capable to perform the detection at run-time. Overall, it can be conclude that the on mote evaluation has shown that the *AID* has achieved the first objective, which is to run on a mote at run-time.

The second objective is to evaluate the degradation caused by on-mote implementation which removes the floating-point number operation. According to the evaluation result shown in Table 3.6, the detection accuracy achieved on a mote without the floating-point number operation does not significantly deviated from the detection accuracy achieved in Matlab simulation which is with the floating-point number operation.

The third objective is to evaluate the overhead caused by integrating the *AID* to the firmware. The result shown in Table 3.7 shows that the *AID* is relatively heavy on the RAM compared to other on-mote resources. Nonetheless, the maximum overhead is only about 20%. The overheads on other on-mote resources are all below 10%. These results mean that the proposed *AID* will be suitable for mote applications.

3.6 Summary

This chapter aims at answering the first research question in this thesis. To review whether it has been answered in this chapter, the research question is shown below for reference:

"How the erroneous usages of ECG device affect the captured signal and whether it is possible to detect the anomalous signal on a mote at run-time?"

To answer this research question, this chapter started based on the finding from Chapter 2. A new algorithm called *AID*, which is proposed to detect the usage

anomalies on a mote at run-time, is used to demonstrate the capability of detecting the usage-related anomalies. Through the Matlab evaluation, evidence has been presented to show that the proposed *AID* is capable to detect the usage anomalies as soon as the signal quality falls below the minimum requirement of a cardiologist. Through the the on-mote evaluation, evidence has also shown that the *AID* is capable to run on a mote at run-time. The on-mote evaluation has also shown that the on-mote implementation will not significantly degrade the detection accuracy of the *AID*. However, through the detail analysis of the evaluation results, there is one drawback for the *AID*. If the heart condition causes the ECG signal become significantly distorted, the *AID* can not always tolerate this case. Although it only happens in some rare cases, it may classify anomalous signal caused by heart conditions to usage anomaly. In the next chapter, the possibility to identify the root cause to the detected usage-related anomaly will be investigated.

Chapter 4

IDENTIFYING THE ROOT CAUSES TO THE USAGE ANOMALIES

The core objective of the research presented in this thesis is to assist the ECG-Based HC-BSNs user to use the device in a correct way so that the system dependability in terms of the availability and the reliability of the captured signal can be assured. In order to assist the user to eliminate the usage issues, a system need to be able to detected the symptom in the captured signals when a usage issue has been made by the user. Then, the system should also has the capability to identify the root causes to the symptom, which is the specific usage issue made by the user. In previous chapter, the possibility of detecting the anomalous signals caused by the erroneous usages for the ECG-Based HC-BSNs has been explored. However, only providing the information of the usage-related anomaly to the user is usually not enough as the anomalies are the symptom but not the root cause. It has been discussed that only using the symptom to search the root causes by a non-professional user may often lead to inaccurate search results [91]. Subsequently, this chapter will focus on identifying the root causes to the usage anomalies. This chapter includes the following contributions:

- **Contribution 1 -** A study to fill the gap where no existing knowledge of the

behaviours of the signals on the ECG-Based HC-BSNs mote has been carried out

- **Contribution 2 -** A methodology which is designed for identifying the root cause to the usage-related anomaly on the ECG-Based HC-BSNs mote without the human intervention is proposed

- **Contribution 3 -** The proposed methodology is evaluated using the real data recorded using ECG-Based HC-BSN mote from volunteers and the result has shown that the root cause to the usage-related anomaly can be identified without human intervention

The rest of this chapter is arranged as follow: In Section 4.1, the existing root cause identification methodologies will be reviewed. Based on the finding from the review, the methodology to identify the root cause to the usage-related anomaly on the ECG-Based HC-BSNs will be proposed in Section 4.2. Then, the detail of how the methodology is proposed will be presented in the subsequent sections including the study to understand the signal behaviour on a mote when usage issue is presented (Section 4.3), the design of the checklist to be used in the methodology (Section 4.4), and how the checklist is automated in the methodology (Section 4.5). The evaluation of the proposed methodology using real data recorded using a ECG-Based HC-BSN mote from volunteers will be presented in Section 4.6. Finally, this chapter will be summarised in Section 4.7.

4.1 Existing works to identify the root cause

The root cause identification has been discussed in many literatures and some methodologies have also been proposed so far. In [61], Nandi et al. has presented a review of identifying the cause of electrical machines when a fault in the machines has been detected. The review has summarised the major faults of electrical machines. Then, the symptoms of each fault are discussed. Some existing methodologies which aim at identifying the cause to each type of faults have also

been reviewed in the paper. In a summary to their review, existing root cause identification methodologies normally involve the following steps:

- Symptom study

  The methodology to identify the root cause to a fault usually start with study the symptom of that fault. The symptoms include various machine running parameters which are constantly monitored by machine control modules. The study compares the running parameters under normal and fault operating states. The outcomes from the study can suggest the parameters' pattern when different faults have occurred and the feature of the patterns. The pattern and its feature can be either in time domain or in frequency domain.

- Symptom feature extraction

  The second step is to extract the specific features according to the outcomes from previous step. For a root cause identification methodology proposed to identify a specific fault, the relevant features in running parameter, which are correlated to the faults, and can be used for the identification, is extracted.

- Symptom feature classification

  A classification method then classifies the extracted feature into different root cause groups according to the symptom study.

Other existing root cause identification methodologies also follow a similar steps despite the different application domains. Some of the existing works also work under similar scenarios to this research although the application domain is different. For example, Yuan et al. [91] has proposed a methodology to identify the root cause to those known problems, for which an operating system does not perform what is required to perform. The scenario they are facing can be summarised as a device or a system does not perform as expected due to the usage issues, and the cause to the problem need to be found and to be provided to the user so that the problem can be fixed. In their work, a trace-based root cause identification methodology has been proposed. Their work also starts with study some existing

scenarios that their methodology need to address. Based on the study, they proposed to use the low-level system behaviour information, which is defined as the sequence of system events, to identify the root cause to a known problem. More specifically, they intend to find the correlations between system behaviours and known problems so that the knowledge of those correlations can later be used to solve new appeared problems. A machine learning classifier can learn the correlations between the system behaviours and the known problems. Once a new problem has occurred, Yuan's methodology will collect the behaviour information incurred during the problem. Then, the behaviour information will be classified based on existing knowledge. Similar methodology can also be seen in [86] and [85] which use the system operating state information is used based on their study to identify the configuration issues to the problems.

For existing ECG devices, similar work, which is used to identify the usage issues causing the anomalous signal in this application domain, can also be found. As discussed in Section 1.2, there are around 4% of usage errors involved in the clinical application of ECG. Subsequently, there should be existing approaches adopted by the cardiologist to address the usage issues. Through the discussion with a cardiologist, it is suggested that many clinical guidelines have been published by authorities in each region or country. In the guidelines, it lists the authorities' standard to the ECG equipments and their set-up parameters. Then, the procedures that the cardiologists or the equipment operators need to follow during the data capturing process. The procedures involves the detail steps from how to prepare the patient, how to place the electrode to the patient and the jobs to do during the recording. How well the guideline is being followed is assessed by an audit, which is described as the recognised way of accessing and improving the practice [21]. More specific, the audit is usually carried out in the form of a checklist [21]. Along with the guideline, there are usually an audit checklist attached which lists various points that need to check when the recording is not as expected. For example, [77] is a clinical guidelines for recording a standard 12-lead electrocardiogram published by The Society For Cardiological Science & Technology. In the checklist, it

checks whether the equipment is set according to standard, whether the operator is trained to standard, whether the electrodes have correctly placed, etc. By carrying out the checklist, it aims at identifying and eliminating the root cause to the usage-related anomalous signals. Such approach to address the usage issues is approved and recommended by The Society For Cardiological Science & Technology. In this methodology, both the guideline and the checklist are designed based on existing literature, and the experience gained through the study and the real world practices. Each item in the checklist can be considered as the features and the checklist can be considered as a way to classify those features. However, it is to the best of the knowledge of the author of this thesis that the above approach is so far only performed by the ECG devices operator or cardiologist manually. None of existing work has put the effort to automate the execution of the checklist to identify the root causes for the usage anomalies of the ECG devices. Also, the existing checklists were developed to be used for diagnosing the usage issues of existing ECG devices. It is not intended to be used for the ECG-Based HC-BSNs. As a result, it may not be suitable for the application scenario of the ECG-Based HC-BSNs. Consequently, although existing methodology adopted by cardiologist is capable to identify the root causes to the usage-related anomalies, it can not be directly used for an ECG-Based HC-BSNs.

4.2   Methodology

Through the review of existing works in identifying the root cause to a fault, a problem, or an anomaly, it is clear that none of existing work will fit the application scenario in this research. Subsequently, a new methodology will need to be proposed to show that the root cause to the usage-related anomaly in an ECG-Based HC-BSNs can be identified on a mote without the intervention. To make it more specific, the following objectives is defined and their challenge is discussed:

Identify the root causes accurately

The first objective is to identify the root causes to the usage-related anomalies, which have been defined in the fault tree in Section 2.1.5, accurately. The iden-

tification accuracy is the fundamental requirement to the methodology. The inaccurate identification will lead to the usage issue not being able to be rectified. The challenge to achieve this objective is the lack of the knowledge between the usage-related anomalies and their root causes in ECG-Based HC-BSNs, which is considered as vital through the review of existing works of root cause identification. To overcome this challenge, this research have to explore the behaviour of the device when usage issues have occurred so that more knowledge can be built for more accurate root cause identification.

Perform the identification on a mote

As discussed in Chapter 3, the frequent wireless communication will drain the onboard battery rapidly compared to standalone operation of the mote. Subsequently, it is also desired that the root identification methodology is capable to be run on a mote without the need to transmit data for off-mote processing. Due to the constrained environment on a mote in terms of the computation ability and memory resources, it significantly limit the methods that can be used.

No human intervention should be needed

The final objective is to identify the root causes without human intervention. As the users of the ECG-Based HC-BSNs are usually untrained non-professionals, it is not expected that they will have the knowledge to diagnosis the usage issue. It is also less likely that there will be a trained professional who will be always available to help the users. Subsequently, it is required that the root causes identification methodology should need no human intervention so that information of rectifying the usage issue can be provided at any time. The challenges to achieve this objective include the lack of knowledge for the device as discussed before, and whether the knowledge can be transfer into a mote executable way especially the identification is desired to be performed on a mote.

4.2.1   Understanding the effects of each root cause

The proposal of a suitable methodology starts with carrying out a qualitative study of understanding the effects of each root cause which have been derived in Sec-

tion 2.1.5. So far, the root causes to the usage-related anomalies are derived based solely on the anomalies appeared in the ECG signals. However, the signal sources on a mote are not limited to ECG signal only. There are many other signal sources depending on the mote configuration. For example, Shimmer2r can also record the motion or activity signals using the on-board three-axis accelerometer. How each of the root cause to the usage-related anomalies in ECG signal can affect other signal sources on a mote is still unknown. For example, it is currently known that when a mote is not attached to the body securely, *Noisy Signal* anomaly can happen to the captured ECG signal. However, whether the insecure attachment can affect the reading from the accelerometer and what the effect to the reading from the accelerometer is still unknown. As a consequence, such qualitative study aims at understanding how each signal on a mote will changed when a root causes to the usage-related anomalies happens. To carry out the study, the real data from the hypothetical monitoring subject will be captured using the ECG-Based HC-BSNs first. When each data sample is being captured, one erroneous usage, which is the root causes to the usage-related anomalies derived in the fault-tree in Section 2.1.5, will be intentionally made. With the erroneous usage being intentionally made to the deployment, the hypothetical monitoring subject will perform some normal activities during the recording process. The captured data sample will be annotated with what erroneous usage is made to the deployment for future reference. Then, the study of each data sample is mainly based on the observation. All the observations are done in time domain. The feature in other domain is less useful as the mote does not have the ability to perform domain transformation due to constrained environment. The data will be plotted and observed for the phenomenon when the usage-related anomaly appeared in the ECG signal. By summarising the phenomena, the outcome from this study is used for the root cause identification. The detail of this study will be presented in Section 4.3.

4.2.2 Design of the root cause identification methodology

In order to achieve the objectives, the high level design of the methodology has been made from the knowledge gained through the qualitative study. The high level design of the methodology is listed as below and justified in the rest of this section:

- Only time domain features is used

- The root cause is identified in conjunction with the signals from accelerometer and the mote's running status (e.g whether the mote is being charged)

- adopting the checklist approach, which is currently used by the cardiologist, to classify the signals' feature

- The identification is carried out by automating the execution of the checklist

As it is desired that the root cause identification can be performed on a mote, the constrained on mote environment will prevent some complex algorithm from being executed. The domain transform is one of those algorithms. The domain transform algorithm (e.g Fourier Transform) requires the capability of performing floating-point operations which is not available on a mote. Even if the transform can be performed, it will generate a huge amount of computational overhead. As a result, the methodology is designed to only use the time domain features to perform the root cause identification.

From the fault tree, it can be seen that some of the root causes is motion related. To detect the motion, the accelerometer, which is widely available on a BSN mote, is the simplest and widely used approach. Subsequently, the aforementioned study of understanding the effects of each root cause has pay more attention to the correlation between signals from accelerometer, usage-related anomalies, and the root cause to that anomaly. The outcomes from the study (detail of the study is presented in Section 4.3) have clearly shown that usage issues can have significant effects on the signals from the on-board accelerometer. Furthermore, the study has

also shown that the mote's running status (e.g whether the mote is being charged) has the correlation to certain type of the root causes. Therefore, the methodology will uses the information of anomaly type in ECG signals, features extracted from the accelerometer, and the mote's running status.

As the root cause identification is a knowledge dependent process, the identification accuracy is depending on how much knowledge has been obtained from previous failures of a system. Therefore, the identification accuracy of the methodology can be improved though iteratively introducing the the new knowledge into the methodology. To make it easier to study the new knowledge gained through the operation of the system, a white-box classification approach (the detail of how the classification is done is visible to the researcher) is more sought than a black-box classification approach (the detail of how the classification is done is not visible to the researcher or understand by the researcher). The aforementioned checklist approach adopted by cardiologist is an white-box approach. A checklist is described as a series of actions, tasks or behaviours which are arranged in a systematic manner [30]. Each action, task or behaviour is checked off when it is completed. After completing the checklist, it is able to identify whether an individual action, task, or behaviour is omitted or is not performed properly. Such approaches are not only used by cardiologists. It has been extensively used in areas like aviation where human error is inevitable [31] especially when the working environment is stressful [74, 35]. It has been proven to be an important tool in error management and performance enhancement, which can make a significant contribution of reducing the risk of costly mistakes [31]. When adopting this approach, every knowledge learnt from studying a system is transferred into one or several checks. When a false identification result is appeared, it can be easily traced down to which check gives the incorrect result and what causes the incorrect result. Subsequently, the modification can be made to the checklist so that further false result under such situation can be avoided. However, as discussed before, the existing checklist may not be suitable for the application scenario of ECG-Based HC-BSNs. That means, in order to use the checklist approach to identify the root causes of the usage-related

anomalies in an ECG-Based HC-BSNs in this research, it is necessary to design a checklist, which is suitable for the ECG-Based HC-BSNs. The detail of how the checklist is design for this methodology will be covered in Section 4.4

In the existing approach of using the checklist to identify the root cause, the checklist execution is carried out by a human. However, it has been discussed and shown in the objective that no human intervention should be needed as it is not feasible to have the checklist executed by a human in the application scenario of the ECG-Based HC-BSNs. As a result, the final design of the methodology is to automate the execution of the checklist so that the root cause can be identified without the human intervention.

Following these high level design, the design of the methodology can be further broken down into the following three parts:

- **Part I:** Understanding The Effects Of Each Erroneous Usage (Presented in Section 4.3)

- **Part II:** Design the checklist (Presented in Section 4.4)

- **Part III:** Automate the checklist execution (Presented in Section 4.5)

During the run-time, the designed methodology includes the following step to identify the root cause to the usage-related anomaly:

- **Step 1:** Anomaly detection for ECG signals
  When the mote starts working, the previously proposed *AID* will continuously perform the anomaly detection against the captured ECG signals.

- **Step 2:** Checklist recall
  Once an anomaly has been detected, the designed methodology will recall the relevant designed checklist to identify the root cause to the detected anomaly. For example, if the *Noisy Signal* anomaly has been detected from the ECG signals, the *Noisy Signal* checklist will be recalled.

- **Step 3:** Feature extraction from relevant sources

  From the recalled checklist, the required features to identify the root cause can be extracted from the captured signals.

- **Step 4:** Feature classification using the recalled checklist

  Classify the extracted features based on the checklist.

4.3 Understanding The Effects Of Each Erroneous Usage

In this section, qualitative study of the root causes to the usage-related anomalies which have been derived in Section 2.1.5 will be presented. The purpose of this study is to understand the effects on the possible signals on a mote when each erroneous usage is made toward the ECG-Based HC-BSNs. To carry out the study, signals will be first recorded from real subjects. The ECG-Based HC-BSNs mote used to record the signals from subjects in this research is the Shimmer2r with ECG daughter board [2]. With this hardware configuration, it is capable to physically record the ECG LEAD-II (RA-LL), ECG LEAD-III (LA-LL), and the acceleration on X-axis, Y-axis and Z-axis. Furthermore, mote running status (e.g whether the mote is being charged) can also be recorded. During the recording process, each of the usage-related anomalies will be triggered according to the fault tree and the signals are recorded. The knowledge from the recorded signals is learnt by observing the feature of the recorded signals under different scenarios.

During the signal recording process, the recording parameters (e.g sampling rate) will need to be unified so that it is meaningful for the signal recorded from difference to be compared and analysed. For the Shimmer2r with ECG daughter board, the configurable parameters include the sampling rate, accelerometer range, and to record the raw or the calibrated data. These configurable parameters can be set via the Multi-ShimmerSync software running on a personal computer before recording the signal. In this research, the following parameters are used. For the sampling rate, $256Hz$ is used as such sampling rate can fulfil the minimum require-

ment to record a useful ECG signal. For the accelerometer range, it is configured to sense the acceleration between plus and minus one and half time of the gravity ($\pm1.5G$, equals to $14.7m/s^2$), which is sufficient to capture the acceleration of normal activities and still achieve good resolution for very tiny amount acceleration changes. As the on-board Analogue-to-Digital Converter (ADC) has maximum of $12\text{-}bit$ resolution, the bigger the accelerometer range, the less resolution for tiny acceleration changes. Finally, the raw data is chosen to be recorded in this research mainly for the reason that the raw value from the ADC is an integer. If calibrated data is used, it will involve the use of floating-point number operation which is unavailable on a mote. As the raw value is used, it is inappropriate to assign any unit for the amplitude readings. Therefore, all the signal plots in the rest of this section will be without unit for the amplitude. Also, due to the sampling resolution being $12\text{-}bit$, the amplitude readings will be between 0 and 4095 with the 0 being $-1.5G$ and 4095 being $+1.5G$. As a consequence, the readings from the accelerometer when the mote is stationary may not be zero.

In terms of the scenarios that need to be studied and understand, this study will mainly focus on the scenarios where either the *Blank Signal* anomaly or *Noisy Signal* anomaly is presented in the signals. The reason for this is that the outcome from the fault tree analysis shows that for the *Inverted Signal* anomaly, the only cause to it is the electrodes' location being swapped. Therefore, no decision will need to be made to identify the root cause of *Inverted Signal* anomaly. However, as the mote being attached to wrong area, according to the fault tree, can lead to either *Blank Signal* anomaly or *Noisy Signal* anomaly, a separate scenario will be study for such root cause. Furthermore, to better understand the scenarios that this research are focusing, it is necessary to study and understand the behaviour of each signals when no usage issue has been made. In the rest of this section, the study for each scenarios will be presented.

4.3.1   Scenario 1: No usage-related issue

As introduced in Section 2.1.3, for the ECG signal to be useful to the cardiologist, the following deployment requirement must be met: the ECG electrode pads

must be securely attached to the correct position; the cables from the electrode pads must be connected to the correct channel on the mote; the cables must be securely fixed to the body and well insulated using medical tape; and the mote must also be securely attached to the body using the medical tape. Figure 4.1 illustrates one of the recorded signal when above criteria has been achieved and the monitoring subject was sitting stationary. Through the study of the recorded signals, it is clear that the ECG signals are free from usage-related anomalies as expected. The accelerometer can detect a tiny amount of acceleration. Although the monitoring subject is stationary as being seen, there are actually some tiny body movement. Such tiny body movement is estimated to be caused by the essential body movement when breath as the mote is deployed on the chest area where body movement caused by breath is quite significant. As the accelerometer is set to sense a narrow range of acceleration, the body movement can be captured by the accelerometer.

The above behaviours were observed from the signals when no usage-related issue has been made and the monitoring subject is stationary. However, for the long-term ECG monitoring application, it is not realistic to expect the monitoring subject remaining stationary for the duration of the monitoring. Therefore, this study has also looked into the case where the monitoring subject performs some daily activities while no usage issues has been made. In theory, when the correct usage criteria has been met, the ECG signal should mostly remain useful although small amount of noise, which is below the limit set by cardiologist, may present in the still useful ECG signal, or some less useful RR-Interval is presented occasionally. Figure 4.2 illustrates a signal example of this case. In this specific example, the signals were recorded when the monitoring subject was brushing teeth. From the signal plots, it can be clearly seen that there is stronger acceleration being recorded by the accelerometer compared to the acceleration in Figure 4.1. Under such daily activity, the ECG signals on both leads are still useful to the cardiologist. However, it can also be seen that occasionally, the noisy may presented in the signal. For example, signal from beginning till around 150th sample on the plot labelled as ECG
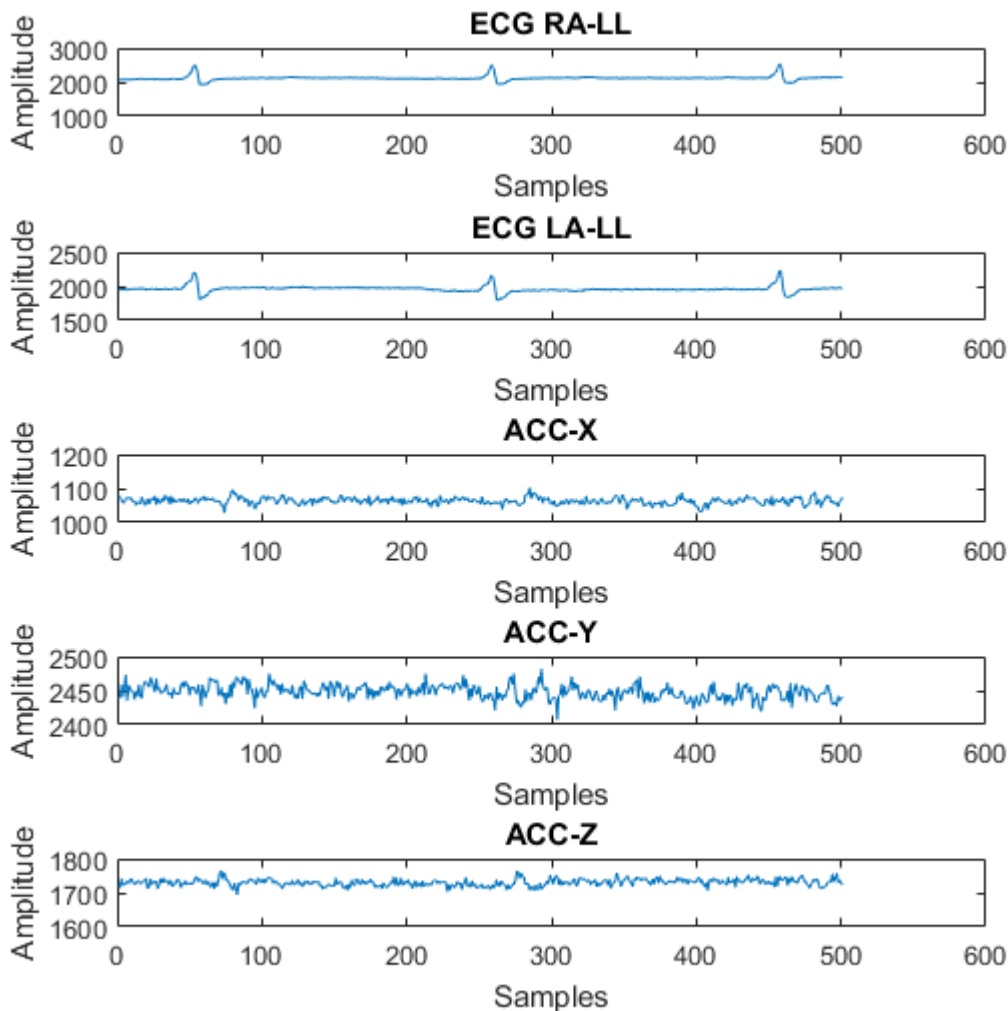
Figure 4.1: A signal illustration when no usage-related issue has been made and the monitoring subject is stationary. From top to bottom, each plot shows the ECG LEAD-III, ECG LEAD-II, X-axis Acceleration, Y-axis Acceleration, and Z-axis Acceleration. The conclusion of this scenario is that when no usage-related issue has been made to the ECG-Based HC-BSNs, the ECG signal is clean and acceleration can be detected from any axis. However, the detected acceleration is very small (in the case shown in above figure, it is less than $0.07G$ if using linear calibration).

LA-LL is noisier compared to the rest signal. During the study, signals, which are recorded when the monitoring subjects were doing other daily activities, have also been observed with similar phenomena. Subsequently, the above phenomena is believed to be applicable for wider range of cases where no usage-related issue is made to the ECG-Based HC-BSNs.

In a short summary to this study, the following features can be observed from the signals in this scenario:

- ECG signal is free from anomaly

- Normal activities will not significantly affect the quality of the ECG signal if the usage criteria has been strictly met

- Even if the monitoring subject is stationary, small acceleration can still be picked up by the accelerometer. Such acceleration is estimated to be caused by the essential body movement when breath

4.3.2   Scenario 2: *Noisy Signal* anomaly is presented

The root causes to the *Noisy Signal* anomaly include the mote being attached to the wrong area, intensive body movement, and insecure device attachment which is then divided into poor mote fixture and poor electrodes insulation. As the mote being attached to the wrong area can also lead to the *Blank Signal* anomaly, it will be studied separately. Subsequently, the rest of this section will look into the case of intensive body movement, electrodes being not well insulated, and mote being insecurely attached to body.

When intensive body movement is presented

When the monitoring subject is performing intensive activities, the rubbing between skin, cloths or other objects will create significant amount of static noise. When such noise is stronger enough, it can still penetrate the insulation of the electrodes, their cables, and the mote. Figure 4.3 illustrates an example when intensive body
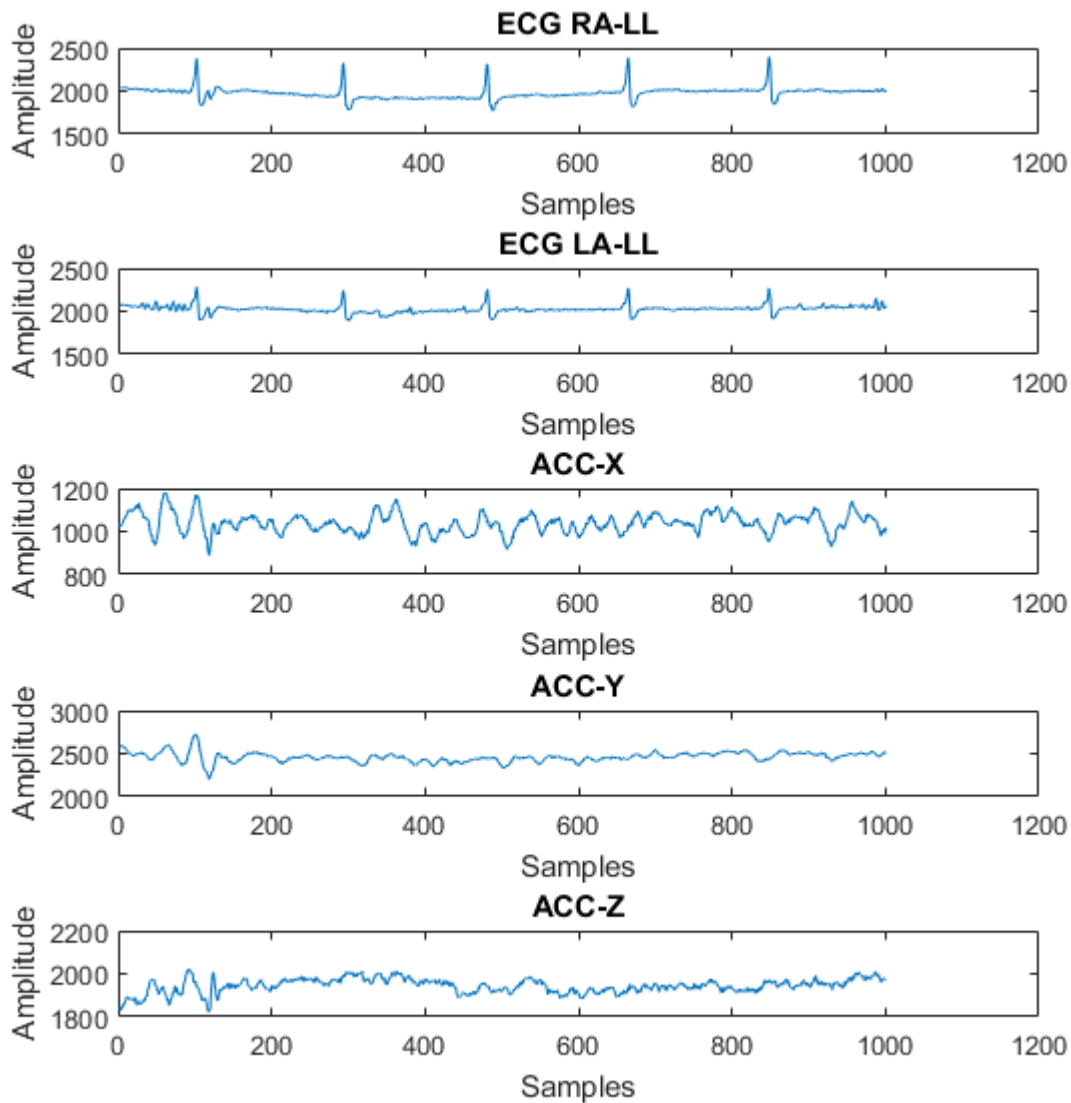
Figure 4.2: A signal illustration when no usage-related issue has been made but the monitoring subject was brushing teeth. From top to bottom, each plot shows the ECG LEAD-III, ECG LEAD-II, X-axis Acceleration, Y-axis Acceleration, and Z-axis Acceleration. The behaviours of these signals support the theory that when the correct usage criteria has been met, most of the ECG signal should still remain useful although some RR-Interval may become less useful occasionally.

movement is presented. It can be seen from the figure that the ECG signals on both leads are no longer useful to the cardiologist. At the same time, dramatic change on the values from the accelerometer can also be seen. Comparing the accelerometer reading from Figure 4.3 to Figure 4.1 and Figure 4.2, it can be seen that the range of acceleration reading in this case is significant wider than both case in Scenario 1. The range of the value from accelerometer in this case is between 0 to 4095, which is the maximum and minimum value a $12 - bit$ integer can represent. On the other hand, the maximum range in Scenario 1 is less than 400 as shown in Figure 4.2.

When electrodes is not well insulated

For the electrodes being not well insulated, the degree of poor insulation can have a significant impact to the outcome of this study. As a result, the signal recording process starts with all usage criteria are met. Then, the insulation material of the electrodes and their cables, which is medical tape, is gradually removed until no insulation materiel is left to insulate the electrodes and their cables from noise. Following such signal recording approach, the following outcomes are summarised.

Firstly, whether the poor insulation of the electrodes and their cables will affect the quality of the ECG signal is depending on the intensity of the activity which is carried out by the monitoring subject. As summarised for the intensive body movement case, the intensive activity will unavoidably lead to the *Noisy Signal* anomaly no matter how well the insulation is. However, on the other hand, the ECG signal's quality will not be affected even if the electrodes and their cables are not insulated as long as the monitoring subject remain stationary.

Secondly, the effect to the ECG signals for different level of insulation is proportional to the intensity of the monitoring subject. In other word, the worse the insulation of the electrodes and their cables, the less activity intensity is needed to make the ECG signal become useless to cardiologist.

Thirdly, if the accelerometer reading is stronger than the reading when monitoring subject is stationary and weaker than the reading when intensive body movement
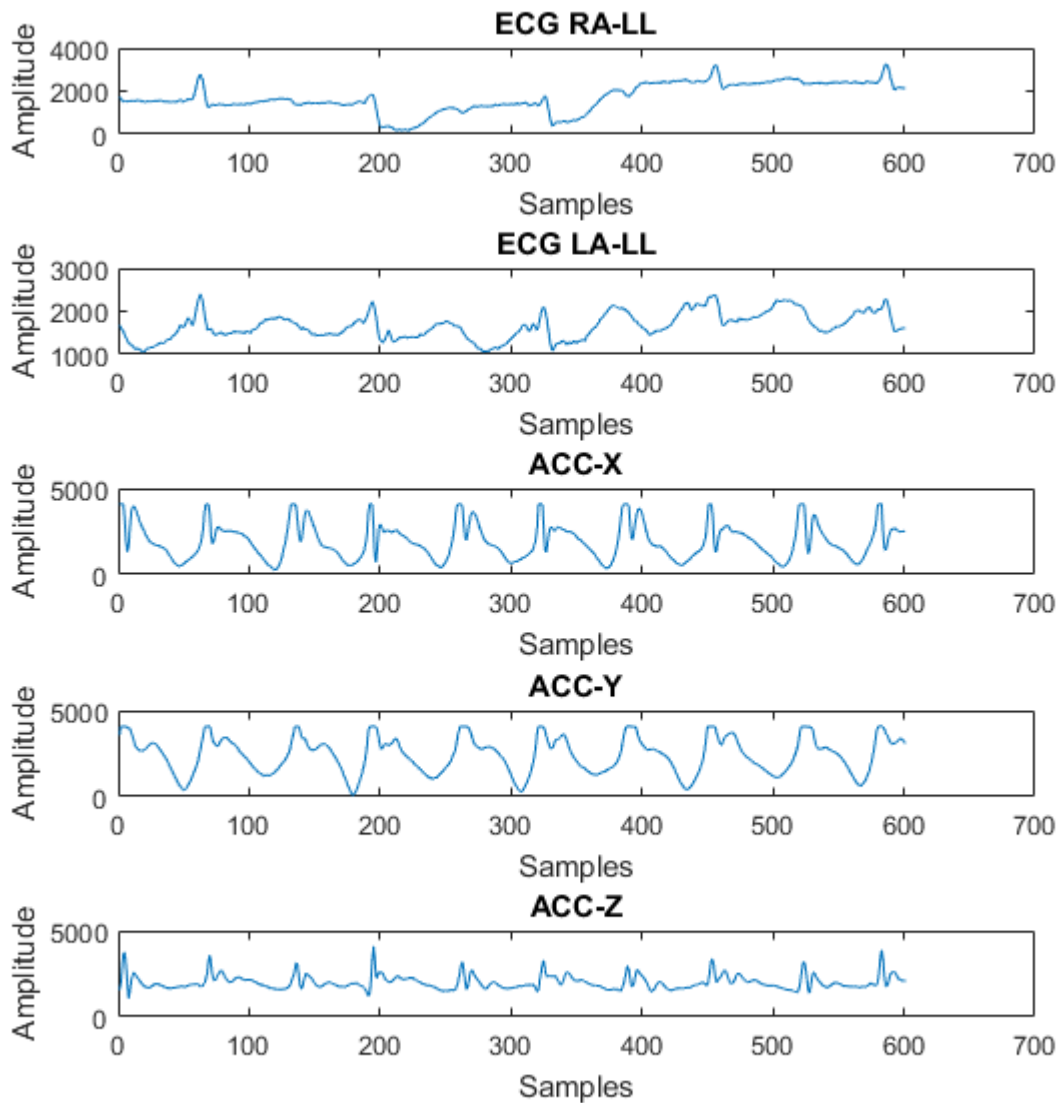
Figure 4.3: A signal illustration when intensive body movement (in this example, the subject is jumping) is presented. From top to bottom, each plot shows the ECG LEAD-III, ECG LEAD-II, X-axis Acceleration, Y-axis Acceleration, and Z-axis Acceleration. It can be seen from the figure that the ECG signals on both leads are no longer useful to the cardiologist. Meanwhile, dramatic change on the values from the accelerometer can also be seen.

is performed, it is possible to decide whether the poor insulation causes the *Noisy Signal* anomaly.

When mote is insecurely attached to body

The signal recording process in this case is also started with all usage criteria are met. Then, the material (in this research, it is also medical tape) used to fix the mote onto monitoring subject's body is removed. The signal is then recorded. Some similar outcomes can be drawn from the recorded signals compared to the case where the electrodes are not well insulated. There are also some unique features in this case.

Similar to the case where the electrodes are not well insulated, how severe the ECG signal's quality being affected is also depending on the activity intensity that the monitoring subject is performing. If the subject is stationary, no effect on the ECG signal will be observed. On the other hand, if intensive activity is performed, the ECG signal will be corrupted by the *Noisy Signal* anomaly unavoidably.

When the monitoring subject is not stationary and normal activities are performed, the following features can be observed from the mote motion pattern and the reading from accelerometer. When a mote is not securely attached to body, it is hanged on the body through the cables, electrodes, and the medical tape. As a result, the mote is flipping on the body. In this case, the mote's orientation is constantly changing. As illustrated in Figure 4.4, the X-axis and the Z-axis readings from the accelerometer is constantly change up and down and remain for a short while. Such phenomena is caused by the gravity making effects on different axis of the accelerometer's reading depending on how the mote is rotated. In other word, the gravity is sensed by different axis of the accelerometer. As the direction of the gravity will never change, therefore, it is the mote's, subsequently, the accelerometer's orientation being changed.

4.3.3   Scenario 3: *Blank Signal* anomaly is presented

According to the derived fault tree, the *Blank Signal* anomaly is caused by either mote being detached from target or the mote being attached to the wrong area.
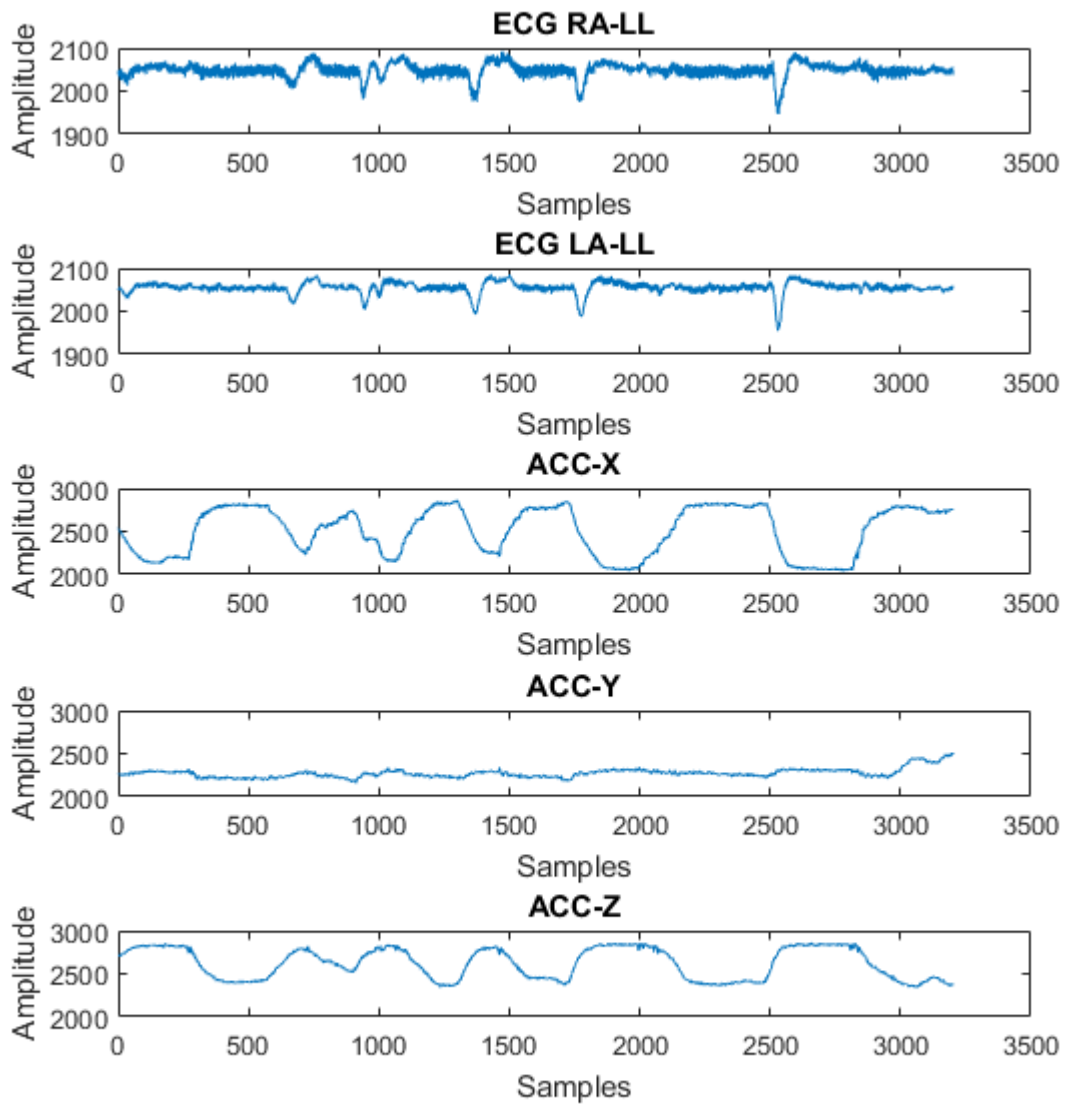
Figure 4.4: A signal illustration when the mote is insecurely attached to body. From top to bottom, each plot shows the ECG LEAD-III, ECG LEAD-II, X-axis Acceleration, Y-axis Acceleration, and Z-axis Acceleration. It can be seen from the figures that the ECG signals on both leads are no longer useful to the cardiologist and the X-axis and the Z-axis readings from the accelerometer is constantly change up and down.

For the mote being detached from target, it can then be divided into the mote being removed for maintenance, mote being accidentally detached, and the user being not cooperate. As the mote being attached to the wrong area can also lead to the *Noisy Signal*, it will be studied separately. This section will mainly focus on the cases where the mote is detached from the monitoring subject.

To study the case where the mote being detached from monitoring subject for maintenance purpose, it is necessary to define what the maintenance means in the application of ECG-Based HC-BSNs. Here in this research, the maintenance is referred to the regular renew of the electrode pads and the recharge of the mote. For the electrode pad, as it is in direct contact to the skin, it is essential to renew for hygienic purpose and for the contacting skin to recover. The mote's battery will also need to be recharged regularly. Furthermore, the battery will need to be recharged more frequently then the electrodes pad will need to be replaced. For example, the mote used in this research will need to be recharged at least once per hour. On the other hand, the manufacturer of the electrode pad used in this research recommends the pad being change every 72 hours. Subsequently, the electrode pads renew and mote battery recharge can be done at the same time. Due to the frequency that a mote need to be recharged, this study will only consider the mote being removed for recharging as the maintenance. To recording the signals during the mote being removed for maintenance, the recording process starts with the mote being attached to the monitoring subject correctly. Then, the deployment will be removed gradually (e.g remove the medical tape one by one) so that the whole mote will be eventually detached from the monitoring subject. Finally, the mote will be plugged into the charger. This sequence is believed to replicate the real process in which way the mote is removed for maintenance. The recorded signals have shown a sequence of change following the process of removal. At the beginning, as the mote is working correctly, the recorded signals are free from usage-related anomaly, which have been shown in Figure 4.1 and Figure 4.2. During the process of removal, as the deployment is gradually removed, the *Noisy Signal* anomaly will be introduced and the noise strength level will be-

come stronger and stronger during this process. Until the mote and the electrodes are fully detached, the recorded ECG signals will change from containing *Noisy Signal* anomaly to the *Blank Signal* anomaly. At last, as the mote is plugged to the charger, the mote running status will show that the mote is being charged. Since this point, the ECG signals will remain as containing *Blank Signal* anomaly and there is no acceleration can be detected by the accelerometer.

When the mote is detached due to the subject is not cooperative, the process of how the mote is detached from the monitoring subject in this case is the same to the the mote being detached for maintenance purpose. The recording process also starts with the mote being attached to the monitoring subject correctly. Then, the deployment will be removed gradually so that the whole mote will be eventually detached from the monitoring subject. After the mote is detached, the mote is left on a surface (e.g desk) without further touching, or the mote is plugged to charger for a longer than necessary time. Following such procedure, the recorded signals have also shown the same sequence of changes as shown in the case where the mote is detached for maintenance. However, in contrast to the mote being detached for maintenance purpose, the key difference is that the mote is not being charged after detachment, or the mote is being charged for a significantly longer period than necessary.

The third case of mote detachment is that the mote is detached accidentally. What it means is that the mote comes off from the monitoring subject without the intentional intervention from the subject. One real example will be the mote or electrodes being detached from the subject during sleeping as the subject may roll over during the sleep, and the force or the squeezing may make the mote or electrodes detach from body. During the signal recording of this case, it is appeared that the detachment does not always mean the whole devices (mote and electrodes) being detached. Instead, most of the cases are that only part of the attachment is detached (e.g one electrode is detached and others are still attached), or become insecure attached. For the later case, it will likely to lead to the *Noisy Signal* anomaly being detected, which has already been covered in Section 4.3.2 so that

it will not be discussed here. When part of the attachment is detached from the subject, the recorded signal has also shown that it requires the subject to remain stationary so that the recorded signal contains *Blank Signal* anomaly. Otherwise, the signal will also contain *Noisy Signal* anomaly instead of *Blank Signal* anomaly. Combining the mote being partially detached and the subject remaining stationary, the following outcomes are concluded for this case. Firstly, the ECG signal from the detached electrodes will contains *Blank Signal* anomaly whereas ECG signal from the still attached electrodes will remain anomaly free. Secondly, as the mote is still being partially attached, acceleration can still be able to detect by the accelerometer. The intensity of the acceleration is similar to the case where no usage-related issue has been made (as shown in Figure 4.1).

4.3.4 Scenario 4: When a mote being attached to the wrong area

Before starting the study of this scenario, it is necessary to clarify the definition of what is meant of wrong area in this research. In the deployment instruction of the Shimmer2r, it has been clearly stated that the mote need to be deployed to the chest area. Therefore, the wrong area is defined in this study as the wrong chest area compared to the location indicated in the deployment instruction. Other locations (e.g deployed to the back or side of a person) will not be considered. For example, the case where the mote is deployed to the chest area lower than shown in the instruction will be considered in this study. However, deploying the mote to the back of a person will not be considered in this study as it is not the chest area. With such clarification, the signals were recorded. The results from this study can be categorised into two groups.

In the first group, the deployment area is not significantly different compared to the instruction. For example, the deployment is $2cm$ below the area where it should be. In this case, the mote can still capture valid ECG signals and other signals are generally the same to the study when no usage-related issue is presented, which has already been illustrated in Figure 4.1 and Figure 4.2.
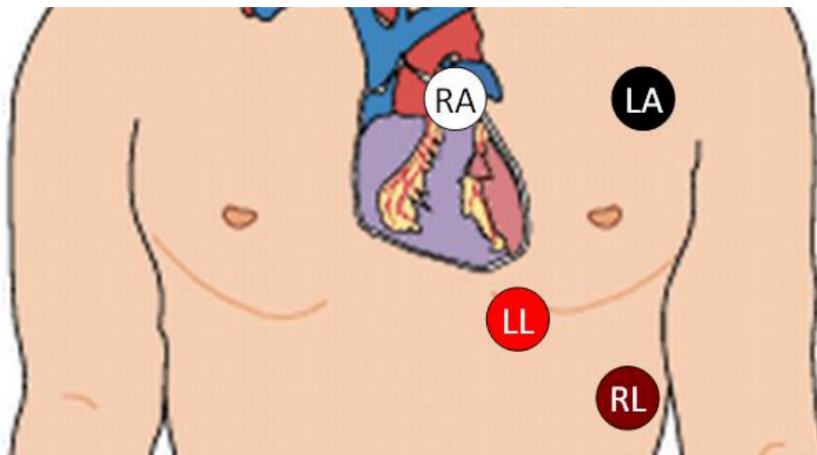
Figure 4.5: The locations that each electrode pads need to be attached to and their connection to the mote's four ports

In contrast to the first group, the signals, which are used to gain the results in the second group, are all recorded when the deployment area is significantly different compared to the instruction. For example, when a person deploy the ECG-Based HC-BSNs to himself/herself according to the electrode location diagram (an example diagram for Shimmer2r is shown in Figure 4.5), one common mistake is to mirror what is shown in the diagram. In this case, due to the mirror effect, the mote is actually deployed to the right chest, while the written instruction says the mote should be deployed to the left chest. When the deployment area is significantly different to the instruction, it can either causes the ECG signals contain *Blank Signal* anomaly or *Noisy Signal* anomaly. Although it can lead to two different types of usage-related anomalies, some common features are shared no matter which anomaly is contained in the ECG signals according to the study of the recorded signals. As this scenario can only happen during the deployment, the recorded ECG signals are neither correct before nor now. If the deployment is always to the wrong area, the system will never enter a correct operation state before the anomaly is detected. For the readings from accelerometer, as the monitoring subjects are trying to deploy the mote to their body, acceleration can be detected as the mote has been attached to the body. The intensity of the acceleration is similar to the case shown in Figure 4.1.

Page 112

4.3.5 Outcomes

Though carrying out this qualitative study of the behaviour of signals available on a mote, the following outcomes can be summarised:

- The anomaly and its root cause is strongly correlated with the behaviour of the signals from accelerometer

- The useful features that can be extracted from the accelerometer reading include whether activity is detected, the level of intensity of the activity, the orientation stability

- The behaviour which describe sequence of how the detected anomalies are changed is also an important feature

- The useful feature from the mote running status is the battery charging flag

4.4 Design the checklist

As summarised by Hales et al. [30], the following aspects will need to be considered when designing a medical checklist to be used in real world:

- **Context** - Where the checklist will be used and who will use the checklist

- **Content** - What items will need to be checked in the checklist

- **Structure** - The sequence in which the items will be checked in the checklist

- **Image** - The user interface design of the checklist

- **Usability** - Is the checklist being realistic to be used in reality

In this research, the checklist is not designed to be used by any people. Instead, it is designed to be executed automatically by the mote in the form of an algorithm. Subsequently, it is not necessary to consider the image and the usability of the

checklist. The main goal of design the cheklist in this research will be the definition of its structure and content. By defining the content and the structure in the checklist, it will also define an unique path from the anomaly to its root cause. As suggest by Heuvel et al. [34] and Rooney et al. [70], defining the path from a system's failure to the causes to that failure involves the use of a decision diagram. In this research, decision tree approach is used as the decision diagram to aid the representation of the checklist's content and structure. Similar to previous section, the main focus of usage anomalies will be the *Blank Signal* anomaly and the *Noisy Signal* anomaly as the *Inverted Signal* anomaly has only one cause to it and no further check is required to identify the root cause.

4.4.1 Decision Tree Analysis for *Blank Signal* anomaly

According to the fault tree, when the *Blank Signal* anomaly is detected, it can be either caused by the mote being attached to the wrong area on human body, or the mote is detached from the monitoring target. To make decision between these two causes, the signal behaviour, which describe the sequence of how the detected anomalies is changing, is used. As shown in Figure 4.6, the decision between these two causes is made by checking whether the signal is correctly presented previously. If the signal was correctly presented previously, it indicates the mote is attached correctly before, and now the attachment is incorrect. That means the mote is likely to be detached from the monitoring target. On the other hand, if the signal was not correctly presented previously, that means the user may be trying to re-deploy the mote for any possible reasons (e.g re-deploy after maintenance) but the user can not deploy the mote correctly, or the mote has been detached from the monitoring target for a while. To justify between these two causes, another decision is needed. Such decision is made by checking whether any activity can be detected from the mote. If no activity can be detected from a mote, it is likely that the mote is not attached onto a human body. If a mote is attached to a human body, the study has shown that even if the monitoring target is sitting on a chair or laying on a surface, there are still some minor activity can be detected. Subsequently, if no activity can be detected from the mote, it is considered as the mote

being detached from the monitoring target in the decision tree. On the other hand, if activity can still be detected, the mote is still being operated by the user. As a consequently, it is decided that the mote is attached to the wrong area on a human body.

For the mote being detached from the monitoring target, according to the fault tree, there are three possible causes for the mote being detached from monitoring target. It can either be the mote being removed for maintenance, mote being accidentally detached from the monitoring target, or the user being not cooperative. As shown in Figure 4.6, the decision will be made by firstly checking whether the detachment is for maintenance. In this case, battery charging flag extracted from the mote running status is used. If the mote is detached from the monitoring target and the mote is being charged, it is considered as the mote is removed for maintenance. On the other hand, if the mote is not being charged, it is considered that the mote is detached without a proper reason. In this case, it can either be that the mote is accidentally detached or the mote is removed on purpose due to the monitoring target does not cooperative. As the mote is attached to a human body using medical tapes and self-adhesive ECG electrode pads, there is a chance that the adhesive is not sufficient to hold the mote so that the mote detaches from human body. For example, the body movement when walking may gradually loosen the mote attachment. Eventually, part of the mote will detach from the body. However, in this case, it is unlikely that the mote will fully detached from human body if it is an accident. That means part of the mote will still be hanging on the monitoring target's body. Thus, the activity should still be able to be detected. On the other hand, if the monitoring target is not cooperative, the monitoring target will fully remove the mote on purpose and leave the mote on some place which is static. Subsequently, no activity can be detected from the mote. So, the decision between whether the mote is accidentally detached from the monitoring target or it is removed on purpose due to user being not cooperative is made by checking whether an activity can be detected from the mote. If activity can still be detected, it is the accidental detachment. Otherwise, it is the user being not cooperative.
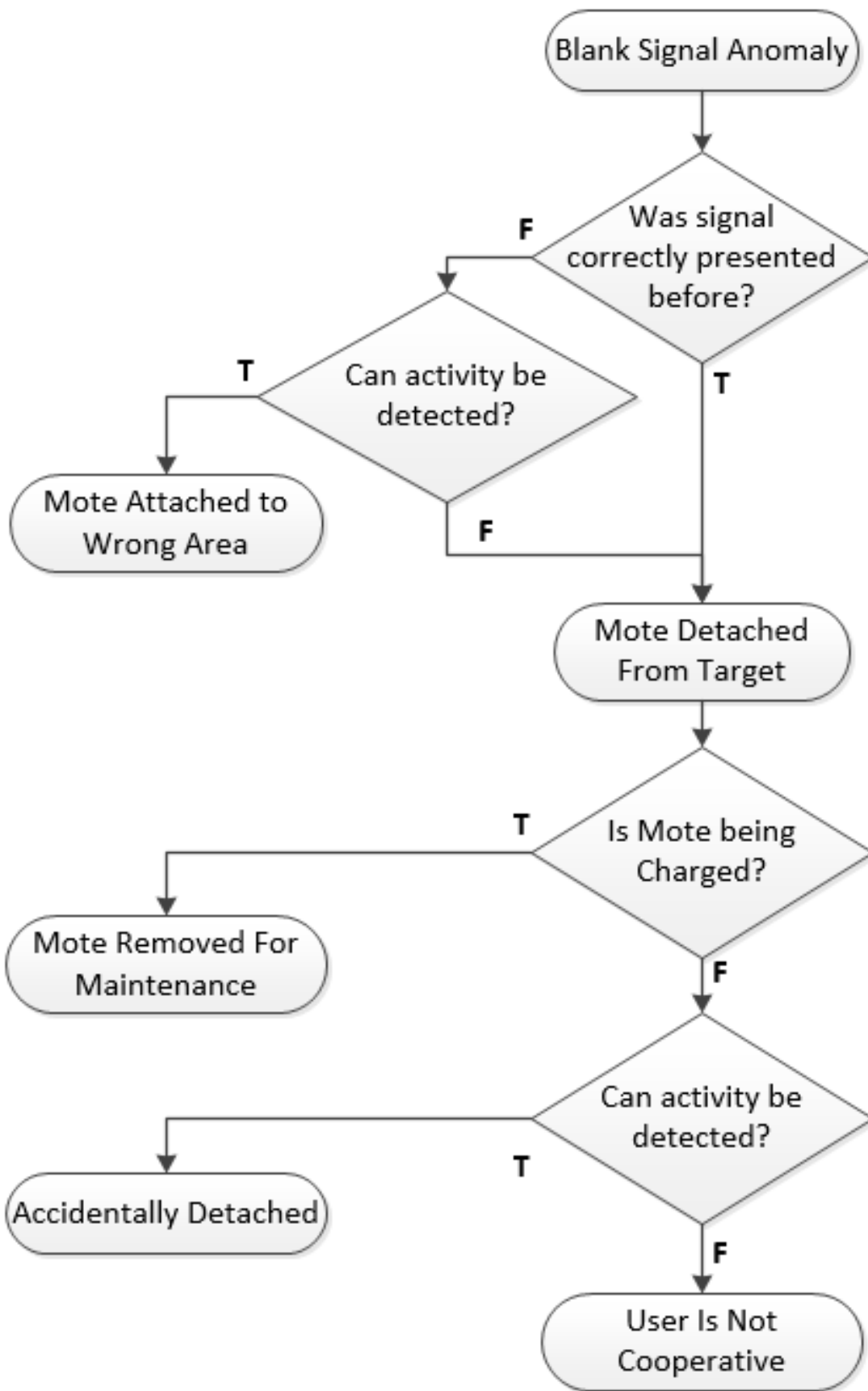
Figure 4.6: Derived Decision Tree for identify the root cause when a *Blank Signal* anomaly has been detected

4.4.2 Decision Tree Analysis for *Noisy Signal* anomaly

According to the fault tree shown in Figure 2.5, there are three causes to the *Noisy Signal* anomaly. It can either be the mote being attached to wrong area, mote being insecurely attached, or intensive body movement being detected. As shown in Figure 4.7, in order to identify which of these causes lead to the *Noisy Signal* anomaly, the first decision is to check whether the signal was correctly presented previously. If the signal was not correctly presented previously, that means the user may be trying to re-deploy the mote for any possible reasons (e.g re-deploy after maintenance or try to rectify the erroneous usage) but the user can not deploy the mote correctly. Subsequently, the root cause to the *Noisy Signal* anomaly is considered as the mote being attached to the wrong area.

On the other hand, if the signal was correctly presented previously and it become anomalous now, it indicates the mote is attached correctly before but may becomes erroneous now or the subject is performing intensive activities. In this case, the only possible erroneous usage is that the mote becomes insecurely attached to the subject. To distinguish between insecure attachment and intensive body movement, it is necessary to check whether any intensive activity has been detected during the period when the signal is anomalous. The detail of how the activity intensity is detected will be covered in Section 4.5.1. If the intensive activity can be detected, it indicates the intensive body movement. If intensive activity is not detected, that means it is more likely to be the mote being insecure attached.

For the insecure attachment, there are two further causes to it. It can be either the mote itself being insecure attached (poor mote fixture) or the electrodes including their cables to the mote being insecure attached. The later one leads to the poor insulation to other noise sources (e.g noise generated by scratching between cloth and skin). To distinguish between these two causes, whether the mote's orientation is stable is checked. The detail of how the mote's orientation is detected will be covered in Section 4.5.1. When the mote attachment is insecure, the mote can flip around the subject's body. On the other hand, if only the electrodes or their cables are insecure attached and the mote is attached securely, the mote's orientation
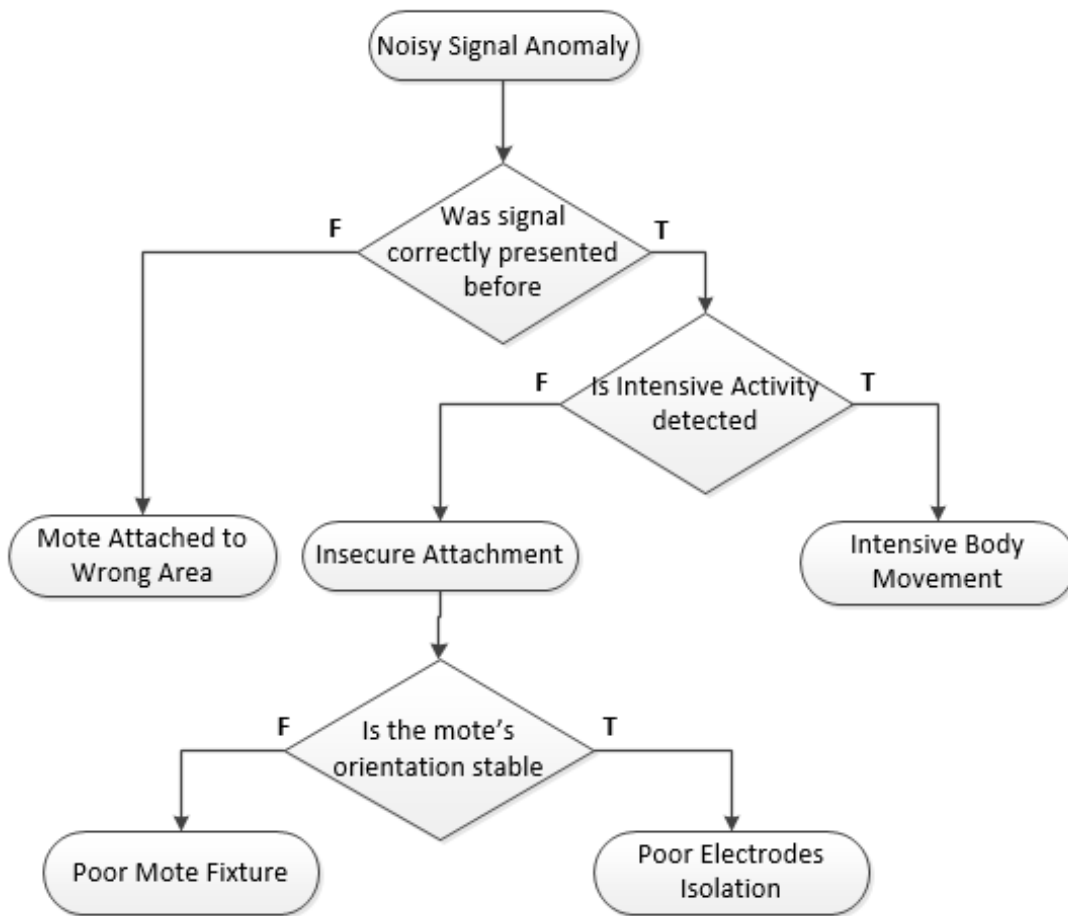
Figure 4.7: Derived Decision Tree for identify the root cause when a *Noisy Signal* anomaly has been detected

will remain stable unless the monitoring subject changes the position (e.g change from sitting to laying down). However, the change of the subject's position will not happen quickly and frequently. For example, it is not likely that the monitoring subject will switch between standing and laying down every second. Subsequently, through checking whether the mote's orientation is stable, it can distinguish between the poor mote fixture and poor electrodes insulation.

4.5 Automated Checklist Execution

With the checklist designed, it is necessary to find a way to execute the checklist automatically by the mote so that the information of the possible erroneous usage that has been made by the user can be prompted to the user. So far, most applications of using the checklist to rule out the human error are still performed by

human themselves. There are many existing work about the electronic checklist technology in nowadays commercial aircraft industry, which is reviewed by Myers in [60]. However, these electronic checklist technologies do not mean the checklist is performed electronically. Instead, these technologies are designed to dynamically and autonomously generate or select the most appropriate checklist under current situation by the system for the users (in commercial aircraft industry, the users are the flight crews) to perform. Subsequently, these works are not fit for the purpose of this research. One existing work presented by Belli et al. [12] has partially automate the execution of the checklist. In their work, they have used the checklist approach for knowledge representation which is later used for programme code review. Their approach to automate the checklist execution is to transform the checks into rules so that an algorithm can follow those rules to perform the check. When a checklist is performed by human, the check can be done using some of the inherent abilities of human like the ability to observe. For example, a human can observe whether the monitoring subject is performing intensive activities. However, a computer does not have such ability. In order to make a computer system to perform similar operation, it is necessary to firstly sense the physical world and convert the event in physical world into digital signal. Then, the signal will need to be interpreted in a way that can be processed by the system. Finally, many rules according to the decision tree will need to be set for the system to perform checks and make decision. From the qualitative study presented in Section 4.3, subsequently, the decision tree, it can be concluded that the readings from the accelerometer are correlated with the root causes to the usage-related anomalies. Therefore, to achieve the automated checklist execution, this research looks into the way to interpret the sensed accelerometer signals, and the rules for the desired checks will also be set. Eventually, the designed checklist will be converted into an executable algorithm. Such algorithm will be capable to be run by the mote and identify the root causes to the usage-related anomalies.

4.5.1 Accelerometer Data Interpretation

In order to make the signals from accelerometer to be useful, this section looks into the interpretation of the signal from the accelerometer. In a review of the wearable accelerometer based motion detectors carried out by Yang et al. [89], the following capabilities of wearable accelerometer based system are listed but not limited to these:

- **Activity recognition -** Based on the signal from the accelerometer which is attached to human body, the activities that the subject is performing can be recognised. For example, Baek et al. [10] proposed an approach to recognise certain types of daily activities. Baek's approach extracts the statistical features from the signal and classify those extracted feature using a multi-layer perceptron classifier. Similar approach can also be found in [67] where Ravi et al. compared many different classifiers to recognise different daily activities using mean, standard deviation, energy, and correlation as the features. The research in this domain mainly focus on the feature extraction and classification approach for the purpose of recognising a specific range of activities.

- **Event detection -**Detecting certain event is with paramount importance in some applications. One example will be the fall detection for the emergent medical rescue for elderly people. Subsequently, many efforts have been put into the research for more accurate event detection based on the wearable accelerometer. For example, Bourke et al. [15] proposed a threshold-based fall detection algorithm and evaluated it over different type of falls like forward falls and backward falls.

- **Estimation of energy consumption -** The daily energy consumption can be useful for maintain a healthy lifestyle. Such energy consumption can be estimated thorough measuring physical activities.

However, through reviewing existing works of using the accelerometer, it appears that existing approaches do not fit for the purpose of this research. Although it

Page 120
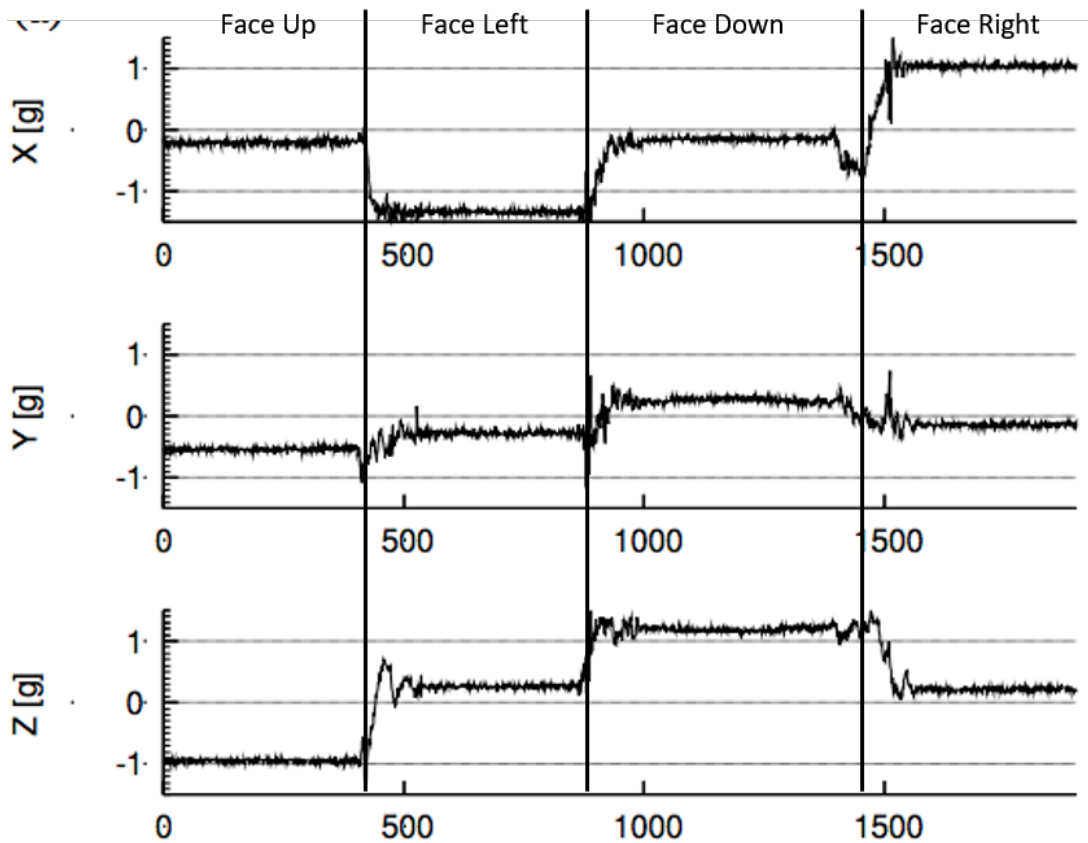
Figure 4.8: Demonstration of the 3-axis accelerometer's readings for each axis when the orientation of the accelerometer is changed (Quoted from [40]. In this example, it records the rolling movement when the subject is laying on bed. When the subject rolls between facing up, facing left, face down, and facing right, the mean value of the accelerometer's X-axis and Z-axis readings for each facing direction is different compared to the other directions. As the subject is only rotating when laying on bed, the mean value of Y-axis reading generally remains unchanged for each directions.

is acknowledged that most existing approaches are more capable than what this research needs, it is not necessary to adopt an overly complex algorithm due to the consideration of resource consumption. Subsequently, a simplified approach based on Ravi's and Baek's approaches is used to interpret the data from the accelerometer. More specific, the mean values calculated from the accelerometer data for each axis are used to interpret the orientation status on each axis whereas the standard deviations of the data for each axis are used to describe the activity intensity on that axis. The effectiveness of each method in terms of the data interpretation will be justified below.

As demonstrated in [40], the mean value of the data from each axis can be used to represent the orientation of the mote. As shown in Figure 4.8, when a mote change its orientation, the mean values of the related axis readings from the accelerometer will change accordingly. In this example, it records the rolling movement when the subject is laying on bed. When the subject rolls between facing up, facing left, face down, and facing right, the mean value of the accelerometer's X-axis and Z-axis readings for each facing direction is different compared to the other directions. As the subject is only rotating when laying on bed, the mean value of Y-axis reading generally remains unchanged for each directions. Such phenomenon is due to the gravity. As the gravity constantly exist and never changes on earth, such phenomenon can be used to represent the mote's orientation. Furthermore, the calculation of the mean value does not require extensive calculation or is not heavy on memory. Subsequently, it is capable to run on a mote at run-time.

To interpret the mote's motion intensity, the standard deviation of the data from the accelerometer is used. The standard deviation is a statistical method that measures the average distance of a data sequence to its mean value. The calculation of standard deviation is shown in Equation 4.1. Figure 4.9 illustrates a period of signal from the accelerometer. This period of signal is captured when the subject is walking, running, climbing, using vacuum, using brush and performing situps, which all have different intensity compared to other activity. The intensity of the signals when each activity is performed can be interpreted by calculating the standard deviation.

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x - \bar{x})^2}{N - 1}} \tag{4.1}$$

where:

s: Standard Deviation

x: Data sequence

N: Total number of data point in data sequence x
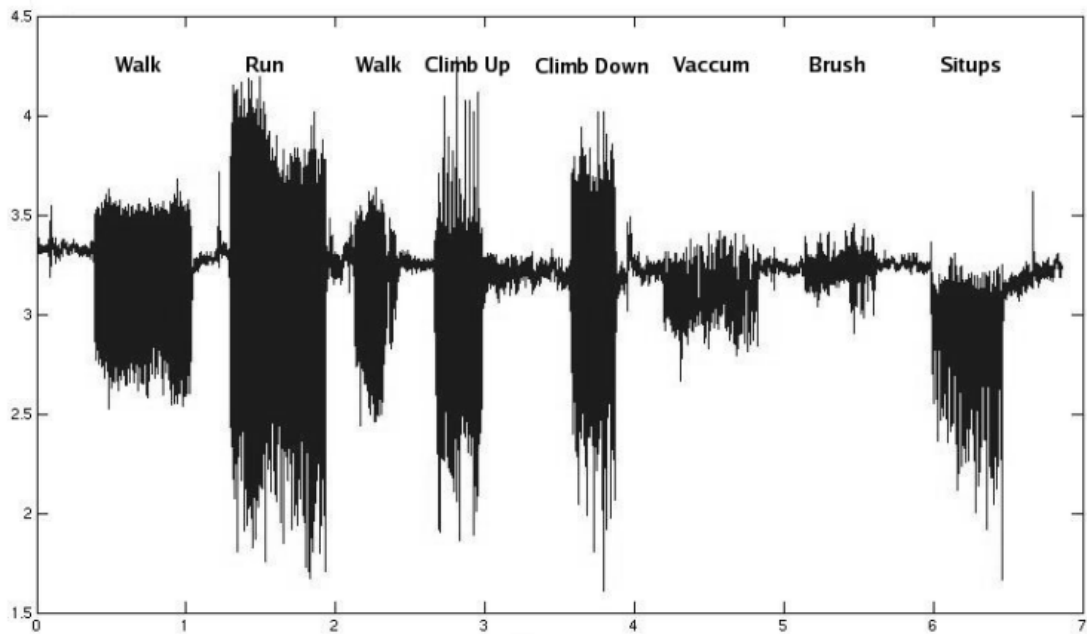
Figure 4.9: Demonstration of the accelerometer's reading when different activities are performed (quoted from [67]). In this example, when each activity is performed, the mean value of the accelerometer's reading remains generally the same. However, depending on the intensity of the activity, the standard deviation of the accelerometer's reading for each activity is different from other.

$\bar{x}$: Mean value of the data x

i: The number of current data point

4.5.2 ACLP: Automatic Check List Performer

Based on the designed checklist and the mote motion status defined using the data from the on-board accelerometer, this section will focus on the implementation of the algorithm called automatic checklist performer (*ACLP*), which automates the execution of the designed checklist. In the designed checklist, there are four checks need to be made in total:

- Can normal activity be detected?

- Can intensive activity be detected?

- Is mote's orientation stable?

- Is the mote being charged?

The first two checks are to decide the mote's current motion status. To make the decision of the mote's current motion status in the designed checklist, a threshold approach is used. The difference between the normal activity and intensive activity in the interpreted accelerometer data is the difference in the calculated standard deviation. To check whether normal activity can be detected, a threshold, T_{ACT}, is defined and used to classify the standard deviation of the accelerometer's reading which is with the same window size of the length of the anomalous RR-Interval. If the standard deviation is less than T_{ACT}, it is considered as no activity can be detected. Otherwise, it is considered as normal activity can be detected. Similar approach is also used to check whether intensive activity can be detected. A threshold, T_{INT}, is defined to classify whether intensive activity can be detected. If the standard deviation is greater than the T_{INT}, it is considered as the intensive activity being detected. Otherwise, the intensive activity is not detected. These two thresholds can be initially set according to the data from the qualitative study. Once it is set, it remains unchanged through out the operation. If further knowledge has gained through the operation. These thresholds can be modified according to the gained knowledge so that more accurate identification result can be achieved.

To check whether the mote's orientation is stable, the *ACLP* continuously records the past five seconds' accelerometer reading. The accelerometer reading is segmented using an one second window with 50% overlap between two windows. Such configuration has been shown to be effective when using the accelerometer data to perform activity recognition in [67, 10]. The mean value is calculated for each windowed reading. Then, to measure how dramatic the mean values has changed, the standard deviation is used. Finally, a threshold, T_{ORI}, is used to classify whether the mote's orientation is stable. If the standard deviation is greater than T_{ORI}, it is considered the orientation is not stable. Otherwise, it is stable.

To check is whether the mote is being charged, there is one specific signal that indicate whether the mote is being charged. Subsequently, this check can be achieved by reading such signal.

Page 124

With each individual check in the decision tree being able to be performed on a mote automatically, the implementation of the checklist is simply the use of branch control in programming (e.g if, switch). The pseudocode of the *ACLP* is shown in Algorithm 2.

Algorithm 2 Pseudocode of ACLP

 1: Pre-set parameters: T_E, $T_{INTENSIVE}$, IS_CHARGING, anomaly_type
 2: **if** anomaly_type is *Inverted Signal* **then**
 3:    Root cause to this type of anomaly is Electrodes Location Swapped
 4: **else if** anomaly_type is *Blank Signal* **then**
 5:    **if** Signal is correct before AND activity is detected **then**
 6:        Root cause is Mote Attached to Wrong Area
 7:    **else**
 8:        **if** IS_CHARGING **then**
 9:           Root cause is Mote Removed for Maintenance
10:        **else if** activity is detected **then**
11:           Root cause is Accidentally Detached
12:        **else**
13:           Root cause is User Being Not Cooperative
14:        **end if**
15:    **end if**
16: **else if** anomaly_type is *Noisy Signal* **then**
17:    **if** Signal is correct before **then**
18:        **if** Intensive activity is detected **then**
19:           Root cause is Intensive Body Movement
20:        **else**
21:           **if** Mote's orientation is stable **then**
22:               Root cause is Poor Electrodes insulation
23:           **else**
24:               Root cause is Poor Mote Fixture
25:           **end if**
26:        **end if**
27:    **else**
28:        Root cause is Mote Attached to Wrong Area
29:    **end if**
30: **end if**

4.6 Evaluation

In this section, the evaluation of the proposed methodology of identifying the root causes of the usage-related anomalies using the real data captured from the voluntary participants in this research will be presented. It aims at demonstrating the evidence of the effectiveness and identification accuracy of the proposed methodology.

4.6.1 Evaluation Metrics

To demonstrate the capability of identifying the root cause to the usage-related anomaly for the proposed methodology, the following metrics are used in this evaluation:

- The number of correct detections by *AID* (Correct Detection)

  As the correct root cause identification relies on the correct usage-related anomaly detection, it is necessary to first record the number of correct detections first. If the number of correct detection is not recorded first, it is meaningless to compare the number of correct root cause identification to the total number of anomalies as the incorrect root causes identification can be caused by the incorrect anomaly detection.

- The number of correct identified root cause of the detected usage-related anomaly (Correct Identification)

  To truly reflect the effectiveness and accuracy of the proposed *ACLP*, the number of correct identified root cause of the detected usage-related anomaly is used. The correct identification requires both the *AID* performing correct detection and the *ACLP* performing correct identification. It is impossible that the *AID* returns false result but the *ACLP* still returns the correct identification result. As a result, it makes more sense to count the number of correctly identified root cause of the correctly detected usage-related anomaly.

- The number of erroneous detections by *AID* (Erroneous Detection)

  In complement to previous two metrics, this metric looks into counting the number of erroneous detections, which is equal to the total number of data sample minus the number of correct detection.

- The number of erroneous identified root cause of the detected usage-related anomaly (Erroneous Identification)

Similar to previous metric, this metric looks into counting the number of erroneous identification. It is defined as the number of correct detection minus the number of correct identification.

4.6.2 Evaluation Setup

The proposed root causes identification algorithm, *ACLP*, has been implemented in Matlab on top of the proposed *AID*. To evaluate the effectiveness of the *ACLP* and its accuracy, test data is fed into the *ACLP* and the result from *AID* and the result from *ACLP* are recorded. Then, the results are compared with a pre-labelled results which are manually done when the test data is captured.

To carry out the evaluation, it requires the data from the ECG sensors and the on-board accelerometer. In this research, the real data captured from the voluntary participants (data acquisition procedure will be introduced in Section 4.6.3) is used instead of the artificially synthesised data. The advantage of using the artificially synthesised data is that it can easily generate a large amount of test data with different anomalies injected, different noise strength and combined with different accelerometer readings. However, the key disadvantage of the artificially synthesised data for this specific evaluation is that the synthesised data may overfit to the combination of the ECG signal and the accelerometer reading to the cases which have been described in the checklist, and the threshold setting. Or, the synthesised data is with the case that never happens in reality. For example, in the decision tree, it is suggested that the *Noisy Signal* anomaly detected by *AID* combined with the standard deviation of the reading from accelerometer being higher than 1 indicates the monitoring target is doing intensive body movement. In this case, when artificially synthesising the test data, it is more likely to use a accelerometer data with standard deviation higher than 1's. Or, the synthesised data may be with the *Noisy Signal* anomaly but the standard deviation of the data from accelerometer is almost 0. In both case, the results will not truly reflect the effectiveness and accuracy of the *ACLP*. Consequently, it is believed that limited number of real data will give better evidence of the effectiveness and identification accuracy achieved by *ACLP* than the large amount of the artificially synthesised data.

4.6.3   Data Acquisition Procedure

The goal of data acquisition process is to record ECG and accelerometer data simultaneously from participants with the real application scenarios of using the ECG-Based HC-BSNs. In this research, Shimmer2r with the ECG Daughter Board [2] is used as the ECG-Based HC-BSNs. To achieve the goal of data acquisition process, the following steps are followed:

- **Step 1 - Experiment discussion with participant**

  The deployment of ECG-based HC-BSNs requires the removal of top cloth so that the chest area is exposed, stick the electrode pads to the body, and fix the device and electrode cables using medical tape. All of these procedures to deploy the device may make the participant uncomfortable or not wish to take part in this experiment. As a result, the first step is to discuss the experiment with the participant. It aims at letting the participant understand what will be going on and making the participant relax. Also, a consent agreement will be signed by the participant by the end of this step to show they are happy for this experiment to go ahead.

- **Step 2 - Participant training** The second step is to teach the participant how to use the device. It simulates the real scenario when the cardiologist prescribes the ECG-Based HC-BSNs to the patient for the first time and teach them how the device is used. More specifically, in this research, the instruction, which will be made available as an instruction sheet for the participant during the entire experiment, for deployment will be as follow. First, the participants will need to take off their top cloth and remove the chest hair where the electrode pads will be attached to. Then, the ECG electrode pads, which are supplied to the participant unused, will need to be attached to the locations as shown in Figure 4.5. Following the attachment of the electrode pads, they will need to be connected to the correct ports on the mote using the supplied cable. The ports on the Shimmer2r's ECG Daughter Board shown in Figure 4.10 will need to be connected to the corresponding electrodes pad shown in Figure 4.5. The final step of deployment is to fix the mote and the

Figure 4.10: The mote with four connection ports with their connection label to each limb cable

cables onto the body using the medical tape, which is also supplied to the participant. The fixture need to be secure for both the mote and the cables.

- **Step 3 - Participant deployment** Following the training, the participant will be asked to deploy the device to themselves in this step. The participant deploy the device according to their understanding of the instruction, and no monitoring of error or advice will be given. This simulates the scenario that the patient deploys the device to themselves at home without cardiologist's instruction or help.

- **Step 4 - Inspection with potential erroneous usage recorded** To obtain the pre-labelled data, in this step, the participant's deployment will be inspected by the research (the author of this thesis). It mainly looks at three points. First, whether the electrode pads are placed to correct location. Second, whether the connection between electrode pads and the ports on mote is correct. Third, whether the whole device is fixed onto body securely.

- **Step 5 - Data recording and annotation** In this step, the actual ECG and accelerometer data will be recorded. During the recording process, the participant can briefly do their daily activities. Also, the participant may be asked to perform some activities (e.g jumping, pushing-up, running etc.) depending on the recorded signal's quality. At the end of the recording process, the participant will be asked to perform some activities which can make the

deployment become erroneous. For example, the participant may scrub the mote deployment area so that the fixture may become loose. This simulates the scenario where the usage become erroneous during the operation of the device which was used correctly at the beginning.

- **Step 6 - Devices removal and consumable disposal** The final step is to remove the device from the participant and disposal of the consumable (e.g electrode pads). The removal is carried out by the participant themselves. The disposal of the consumable can be done by the researcher or by the participant depending on the preference of the participant.

4.6.4   Evaluation Result

Following the data acquisition procedure, there are 130 data samples that had been collected in this research. These data samples were collected from ten voluntary participants who were invited to join this research in person by the author of this thesis. Due to the ethical requirements, all of the participants are male adults. Due to the same reason, the health status of these participants were not asked or assessed before, during, or after this research. Among these 130 data samples, it can be categorised into the following three scenarios:

- **Scenario 1 - No Usage Anomaly**

- **Scenario 2 - Deployment Related Error**

- **Scenario 3 - Usage Becomes Erroneous During Operation**

The evaluation results for each scenario have been summarised in Table 4.1.

For the Scenario 1, there are ten data samples that have been collected during the experiment. Among the ten captured signals, nine of them have been correctly classified as correct signal with no usage-related anomaly. There is one signal that has been classified as containing *Noisy Signal* anomaly by the *AID*. The *ACLP* identified that the root cause to the anomaly is the poor electrodes insulation.

Scenario	Collected Data Samples	Correct Detection	Correct Identification	Erroneous Detection	Erroneous Identification
1	10	9	9	1	0
2	100	100	89	0	11
3	20	20	17	0	3

Table 4.1: The evaluation results for each scenario using the experiment collected data

The Scenario 2 is with the most interest in this research. Thus, more efforts have been spent in collecting data in this scenario. In total, a hundred data samples have been collected during the experiment. Among the a hundred collected signals containing usage-related anomalies, all of them are correctly detected by the *AID*. Then, there are eighty-nine of the detected usage anomalies whose root causes are correctly identified by the *ACLP*. That means there are eleven root causes identifications are incorrect. There is one identification where the accelerometer reports intensive activity has been detected. Subsequently, the *ACLP* identifies the root cause as the intensive body movement which override the true root cause of poor electrode insulation. By comparing the other ten incorrect identified results to the annotated results from the data acquisition process, three out of ten incorrect results are due to the *ACLP* is confused between the poor electrodes insulation and the poor mote fixture. Through checking the signals and the parameters used for each decision in the decision tree, it appears that the parameter used to make the decision is at the edge between the mote's orientation being stable and not stable. As these two root causes are both belong to the insecure attachment of the any part of the whole device, the correction information in the right direction (e.g check attachment security) can still be provided to the user. For the other seven incorrect results, the *ACLP* reports the root cause is poor electrode insulation whereas the annotations from data acquisition process indicate that there should be no electrode insulation issue, but the monitoring subjects were carrying out some activities which are not motion intensive enough to cause noise. However, to perform these activities actually requires quite strong muscle power (e.g push up). A further study for the signals under such circumstance will be presented later.

For the Scenario 3, twenty data samples have been collected during the experi-

ment. The *AID* has correctly detected all of those data samples. Then, the *ACLP* has correctly identified seventeen out of twenty detected usage-related anomalies. However, there are still three anomalies that are not correctly identified. Compared the results from the *ACLP* with the annotations from the experiment, all three wrong identifications are in the case where the *ACLP* is confused between poor mote fixture and poor electrodes insulation.

4.6.5 Further Study of The Incorrect Identification

The evaluation result from Scenario 2 has indicated that under specific circumstance, the *ACLP* does not perform as expected. The analysis of the circumstance leads to a possibility of the ECG signal is corrupted by the Electromyography (EMG) signal. The EMG signal is also an bio-electrical signal on the skin generated by the muscle activity. The EMG signal can be captured by a device using similar principal when capturing ECG signal. As a consequence, it is not rare that ECG and EMG can corrupted each other. It is the nature of these two signals so it is not classified as usage-related issue. There are many existing works which have looked into removing ECG signal from EMG signal [50] or vice versa [69] using different filtering techniques. However, to the best of the author's knowledge, there is no existing work that looks into which specific muscle can corrupt the ECG signal. As a result, a further qualitative study has been carried out aiming at providing information for future research in this problem.

To carry out the study, the mote will be attached to a volunteer without any usage issue. Then, the volunteer will be asked to exercise a specific area of muscles. For example, when measuring whether exercising the muscles around right shoulder area can corrupt the ECG signal, the volunteer will be asked to lift and put down only the right arm so that only the muscles at that area are exercised. Furthermore, as the level of power which the muscle uses can also play an important role to the result, different level of free weights are used to simulate different power output from the muscles. As this study is a qualitative study, the results are mainly based on observing the captured ECG signals for corruption. Overall, the study is carried out with the muscles being divided into five areas: leg area, abdominal area, top

left area, top right area, and neck. For the weight applied to each area of muscle except neck, five different levels have also been applied: no added weight, $2.5KG$ added weight, $5KG$ added weight, $7.5KG$ added weight, and $10KG$ added weight. For the neck muscle, no added weight is applied.

The result of this study shows that the significant effect on the ECG LEAD-III can be observed when exercising the muscles at top left area. With the added weight increase, the effect also become more significant. No effect is observed on the ECG LEAD-II. When exercising abdominal muscle, both LEAD-II and LEAD-III can suffer from muscle noise. When exercising the neck muscle, effect on the ECG signal can also be observed. However, through closer and more detail observation, it seems that the effect is caused by the skin stretching due to the head movement, which subsequently leads to the electrode movement. Such movement seems to be unavoidable due to the stretching force is greater than the force that the electrode adhesive can provide. For the case where other region of muscles are exercising, no effect has been observed from either of the two LEADs. As the ECG LEAD-I on Shimmer2r is calculated from LEAD-II and LEAD-III, it is known whether the LEAD-I will suffer from muscle noise if it is physically captured by the device.

4.6.6 Discussion

Through the limited number of results, it can be seen that for most cases the proposed *ACLP* is capable to correctly identify the root cause to the usage-related anomaly which is detected by the *AID*. However, as the *AID* can not detect and recognise the usage-related anomaly to the accuracy of a hundred percent, there are some incorrect identification as a consequence to the incorrect detection. Subsequently, these results have also shown the importance of the anomaly detection accuracy. The improvement to the anomaly detection accuracy is outside the scope of this chapter, so it will not be discussed here.

For those scenarios where the *AID* reports the correct usage-related anomaly but the *ACLP* identifies the wrong root cause, the results suggest the better tuning of the decision parameters (e.g threshold to decide the intensive activity being

detected). The turning based on existing data can make the result become better. However, due to the lack of data, it is not able to show whether the turning can make real difference in real application of the *ACLP*. For the rest erroneous identification, the extra analysis has suggest the possibility of being corrupted by EMG signal. Subsequently, further study has been carried out and the relationship between muscles and the signals on each ECG channel has been briefly studied. However, the study has to be suspended at this point mainly due to lack of equipment to carry out further study which may also be outside the scope of this chapter.

4.7 Summary

This chapter aims at demonstrating the evidence which can be used to address the second research question in this thesis, which has been quoted below for reference:

> "What is the relationship between the erroneous usages and the signals on a mote environment and whether the root cause to the usage-related anomaly can be identified using the signals on a mote environment and without human intervention?"

To address this research question, a systematic approach has been taken. Firstly, a qualitative study of the captured signal has first been carried out. Such study has shown how the usage issue can affect the signals on an ECG-Based HC-BSNs mote especially the effect on the signals from accelerometer. Secondly, we have designed a checklist to identify the root cause to the usage-related anomaly based on the finding from the study. Finally, we have automated the execution of the designed checklist so that the root cause to the usage-related anomaly can be identified without human intervention. The proposed methodology has been evaluated using the real data which was captured from voluntary participants during this research. The evaluation results have shown its capability of identifying the root cause without human intervention. The evaluation has also looked into the cases in detail where the root cause identification accuracy are suffered. It has also raised the issue where the root cause identification accuracy is heavily depending

on the usage-related anomaly detection accuracy. Subsequently, next chapter will look into the possibility of using off-mote collaboration to improve the usage-related anomaly detection accuracy.

Chapter 5

USING OFF-MOTE COLLABORATION TO IMPROVE DETECTION ACCURACY

In Chapter 3, an on-mote usage-related anomaly detection algorithm, *AID*, has been proposed and evaluated. Through the evaluation, the results have shown that there are some limitations which can affect the detection accuracy under certain scenarios (e.g when the signal is distorted by heart condition). This chapter aims at investigating the possibility of adopting off-mote collaboration to overcome the limitations. Subsequently, the detection accuracy can be further improved. To start the investigation, this chapter will start with analysing the problems with the previously proposed approach. By doing this, the research presented in this chapter is motivated and the possible improvement will also be proposed. In the next section, a literature review of using machine learning approaches to classify the ECG signals will be presented.In the next section, the methodology named *FF-NAID* will be discussed in detail. Then, the evaluation, which looks into various aspects of off-mote collaboration, will be presented. In the last section, the work in this chapter will be summarised.

5.1 Motivation

As introduce in Chapter 1, the ECG-based HC-BSNs is composed of many motes which can wirelessly communicate to each other, or to other devices or servers via a data aggregator. For the proposed *AID*, due to the consideration of energy efficiency, the *AID* is designed to be run on a mote without the communication. However, as the evaluation of the *AID* has shown, although the *AID* can detect the usage-related anomalies with a significantly higher accuracy than existing approaches, there are still some scenarios for which the detection accuracy is suffered. Furthermore, the evidence from the evaluation of the proposed root causes identification, *ACLP*, in Chapter 4 has shown that it can correctly identify the causes to the usage-related anomalies subject to the correct detection from the *AID*. Although the scale of the evaluation of *ACLP* is not large enough to fully cover all possible real life scenarios, it can still show that the bottleneck of the whole idea of assisting the patient to use the devices correctly though indicating the user with erroneous usage when it happen and instructing the user to correct the erroneous usage is the usage-related anomaly detection. Subsequently, the research, which aims at improving the usage-related anomaly detection accuracy, is motivated. The basic idea of the improvement is to adopt the off-mote collaboration so that the on-mote detection can be more accurate. However, as introduced that the communication can consume many times more energy than computation, and the limited on-board resources, it is important to find a way that can balance the cost of communication, on-board resource consumption, and the detection accuracy. In order to achieve this, the rest of this section will firstly formulate the problem that this research is facing. Then, the concept of the improvement will be drafted so that this research is motivated.

5.1.1 Problem Formulation

The key objective of this research is to detect the usage-related anomalies and identify the root causes to those usage-related anomalies. Consequently, it is expected that the algorithm only reports the abnormal ECG signal caused by the erroneous usage as the anomaly. Other abnormal ECG signal caused by none

usage-related reasons should not be reported by the algorithm as the anomaly. For example, Figure 5.1 illustrates four ECG RR-Intervals. The signal illustrated in Figure 5.1a is the correct RR-Interval which is set by cardiologist. Signals in Figure 5.1b, Figure 5.1c, and Figure 5.1d have shown three abnormal RR-Intervals which are caused by different types of heart conditions. Therefore, in the perspective of this research, this signal should be classified as anomaly free as it the abnormality is caused by heart condition instead of erroneous usage. However, if these signals are fed to *AID*, *AID* reports that these signals contain *Noisy Signal* anomaly as the correlation coefficients between these signals to the correct RR-Interval is 0.26, 0.52, and 0.25 respectively. Subsequently, false positive results are reported. There are many different type of heart conditions that can lead to the captured ECG signal being distort from the correct signal. The detail of how each heart condition changing the appearance of the captured ECG signal is outside the scope of this research. Thus, it will not be covered in this thesis. Nevertheless, the key problem that causes the *AID* to report many false positive results is the possible time domain appearance variation of ECG signal even if it contains no erroneous usage.

To overcome this issue, two obvious approaches can be directly adopted to *AID*. First approach is to lower the threshold setting when deploying the *AID* so that more tolerance can be given to the ECG signal containing heart conditions. The second approach is to add more signal templates so that the possible ECG signal variation caused by heart conditions can be included to the templates. However, these two approaches are not feasible to the purpose of this research.

For the first approach, lowering the threshold T_C can give more tolerance to the heart conditions. At the meantime, it can also give more tolerance to the usage-related anomalies especially the *Noisy Signal* anomaly. As introduced in Section 3.3.5, the threshold, T_C, is set by cardiologist by creating noise purposely so that the signal that meets the cardiologist's minimum quality requirement is picked and used to generate the threshold setting. As the threshold is set to the cardiologist's minimum signal quality requirement, lowering the threshold means the
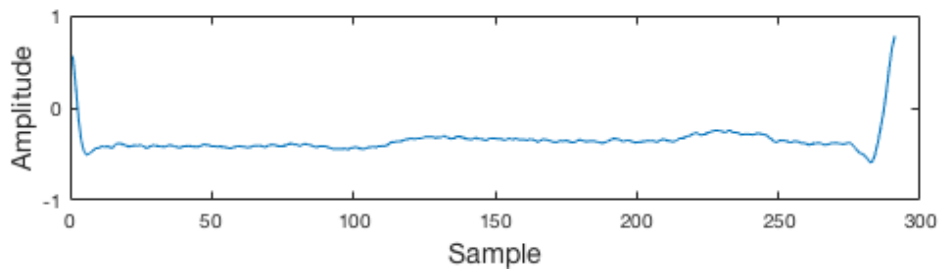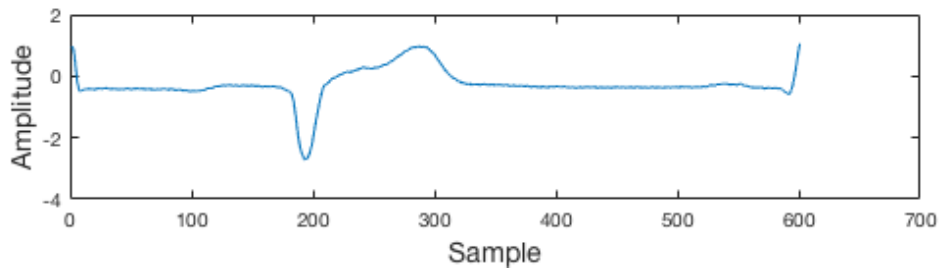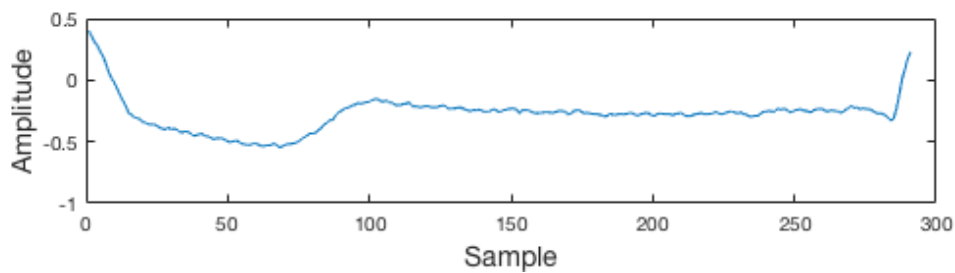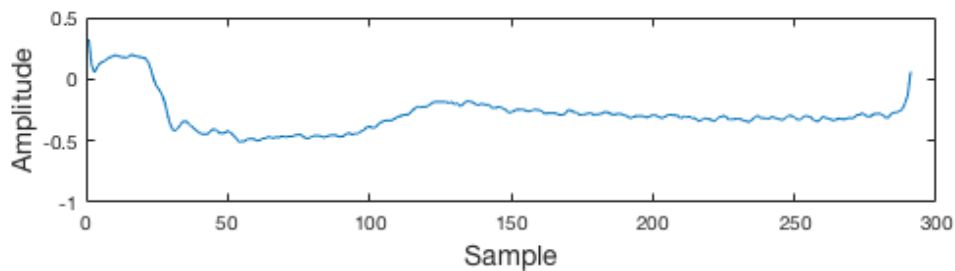
Figure 5.1: An Illustration of some abnormal signals caused by heart conditions. 5.1a shows an example a correct RR-Interval. 5.1b shows first example of abnormal RR-Interval caused by heart condition whose correlation coefficient to the correct signal is 0.26. 5.1c shows second example of abnormal RR-Interval caused by heart condition whose correlation coefficient to the correct signal is 0.52. 5.1d shows third example of abnormal RR-Interval caused by heart condition whose correlation coefficient to the correct signal is 0.25.

anomaly free signals declared by *AID* can still below the cardiologist's minimum signal quality requirement. As a result, this approach is not a feasible solution.

For the second approach, more correct signal templates can cover more normal variation in the perspective of this research so that the classification can be more accurate. However, two factors that make this approach be not feasible for the HC-BSNs. As the *AID* is designed to be run on a mote at run-time without any the need of communication. That means if this approach is adopted, more templates need to be stored on a mote for the algorithm to use. However, the on mote memory, both ROM and RAM, is not sufficient to store all the possible templates. According to a quick test during the research, it is measured that the maximum number of templates that can be stored on a Shimmer2r is around 10 assuming each template is one second in length and sampled at $360Hz$ with 12-bit resolution. Such number of stored templates is much less than it is needed for this approach to be adopted. Although some of the mote platforms have equipped with on-board flash memory, the speed of data reading is not sufficient to support run-time anomaly detection. The computation ability for a mote is another reason that this approach is not feasible for the current *AID* configuration. When each new RR-Interval is captured, the *AID* will need to compare the new RR-Interval to every stored templates. Subsequently, there is not enough time for the mote to finish the classification before the next RR-Interval become ready for detection.

To tolerate the abnormal ECG signal caused by heart conditions while still keeping the same quality level, modification will need to be made to *AID*. In the design of the *AID*, a three stages approach is adopted. The first two stages aim at align the signals and match the signals' size. These two stages are categorised into pre-processing phase. The actual decision of whether the signal is a normal signal or an anomaly is done in the third stage. It uses the Pearson's Correlation Coefficient between **New RR-Interval** and **Predicted RR-Interval**, and a threshold, T_C, to perform the classification, which is believed to be the bottleneck of the detection accuracy. Thus, it is believed that by changing the classification method, the detection accuracy can be improved.

5.1.2 Proposal of Improvement

In order to tolerate the abnormal ECG signal caused by heart conditions while still keeping the same quality level, it is proposed that if the classification approach in *AID* is replaced by a machine learning classification approach, the detection accuracy can be further improved. Meanwhile, by balancing the on-mote computation, data transmission requirement, and the off-mote processing, the overhead in terms of computation, memory consumption, and communication can still be acceptable. The concept behind the improvement is as follow. If the new algorithm can learn many templates including the normal and abnormal templates and a model is built. Such model is usually represented in the form of parameters which is significantly smaller in size than the total size of the templates which the algorithm need to learn. Then, the on-mote anomaly detection can then use the built model to perform the classification. During the run-time, the learnt model can be periodically updated based on the new information which is gathered during the operation. The learning process can be performed off-mote so that there is no need to concern about the constrained on-mote environment when designing the learning process. The new model is then be send to a mote to perform future on-mote anomaly detection. On top of these, as the mote will need to be removed for periodic recharge, the data communications (e.g transmission of new information for off-mote processing) can take place when the mote is being charged. Thus, the communication overhead can be ignored as the mote is powered by external power source instead of on-board battery. Following such process, it is expected that the detection accuracy can be significantly improved in a long term as more and more information is gathered from real world application. Meanwhile, the communication required to send new information and to receive updated model can be reduced in a long term as the information covered by the model is getting more and more comprehensive.

5.1.3 Objectives

Based on the problem formulated in Section 5.1.1 and the proposed improvement to overcome the problem in Section 5.1.2, the following objectives are defined for the research in this chapter:

- **Objective 1 -** Define a methodology based on existing *AID*, which uses the machine learning classifier to perform the detection

- **Objective 2 -** Demonstrate the detection accuracy improvement of the defined methodology

- **Objective 3 -** Demonstrate the effects to the detection accuracy with different configurations to the defined methodology

- **Objective 4 -** Demonstrate the necessity of data sharing between different subjects

The first objective for the research in this chapter is to define a new methodology based on existing *AID*, which uses the machine learning classifier to perform the detection. As it is based on the existing *AID*, the new methodology will also perform the anomaly detection against each RR-Interval through classifying the raw signal. The key challenge here is to find and introduce a machine learning classifier, which can work with the raw signal, to existing *AID* as most of existing approach is worked with features instead of raw signals.

The second objective is to demonstrate the detection accuracy improvement of the defined methodology over the existing *AID*. The challenge here is to find a proper configuration to the defined methodology as there are usually many different configurations can be made to the machine learning classifier. It has also been concluded that a proper configured machine learning classifier is more important than the choice of classifier [41]. Subsequently, it leads to the third objective, which is to demonstrate the effects to the detection accuracy with different configurations to the defined methodology. Such demonstration can also give the evidence where finding the the trade-off between better detection accuracy and energy efficient can be based on.

The last objective is to demonstrate the necessity of data sharing between different subjects. So far, the training data and the classifier is from and used on

the same mote. However, a body sensor network contains many motes. It is not known whether the data from other motes will have any benefit to the on mote detection accuracy. Furthermore, recent literatures have raised the concern of privacy issue[88, 47, 81, 4]. According to those works, many users worry about their privacy due to the information sharing as the medical or physiological signals are private and confidential. By demonstrating the necessity of data sharing between different subjects, it can provide the evidence of whether it is worth to research into the privacy protection if data share is necessary.

5.2   Machine Learning in Electrocardiogram

A widely quoted and formal definition of machine learning is from Tom Mitchell [56], who defines machine learning as below:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E"

The definition of machine learning may sound complicated. However, machine learning algorithms have been widely used so far. It has been used from medical diagnosis, speech, and handwriting recognition to automated trading and film recommendations. It can also be used to make critical business and life decisions every moment of the day. Due to the widely interesting applications of machine learning, many techniques have been developed so far for different applications solving different types of problems. So far, there are many works using machine learning approach to classify ECG signals. In the rest of this section, a review of existing works, which use machine learning in classification for ECG signal, will be presented.

In a survey carried out by Jambukia et al. [38] in 2015 and Salem et al. [73] in 2009, they have reviewed a wide variety of ECG signal classification using machine learning approach. Through summarising their reviews, it can be seen that most of existing works use the following two machine learning classifiers: Support Vector

Machine (SVM) based approach, and Neural Network based approach. In existing literatures, these two approaches can achieved similar classification accuracy. However, the neural network approach is usually criticised due to its computational complexity during training [55]. Also, as how the decision is made by the neural network is hardly understand by the user so that it is sometimes hard to justify the result from the neural network. Figure 5.2 illustrates an example of how SVM can classify the input data (x and y in the figure) into two classes. Basically, the SVM will first find the hyperplane (the dotted line in Figure 5.2) which separates the two classes. Then, the maximum margin from the hyperplane to each data class will also be found (the solid lines in Figure 5.2). When using such model to perform classification, the distance between the hyperplane and the new input will be calculated and compared with the margin to each class. If the distance is above or below the margin on each side of the hyperplane, the input data will belong to that class. On the other hand, for the same problem as the example, if neural network is used, the sum of product between the inputs and weights will be calculated and the result will be propagated to the next layer until the output layer. The output from the output neuron will decide which class the input belongs to. Compared the classification process of the neural network with the SVM, it is clear that the calculations in neural network does not carry any useful meanings to the users whereas the calculations in the SVM can be fully understand by the users. Thus, more and more recent works tend to use the SVM approach.

As conclude in [55], there is no definite answer to which classifier is significantly better than other. The performance that can be achieved by different machine learning classifiers is heavily depending on how the data is processed and its features being extracted ahead of being fed to the classifier. For existing works, none of them feed the raw data to the classifier for classification. Instead, different features, depending on the purpose, are extracted and fed to the classifier. As existing works looked into detecting those abnormal signals caused by heart conditions, the measurement of each ECG components are usually used as the feature. For example, Vijayavanan et al. [84] used the P, Q, R, S, T peak points, QRS com-
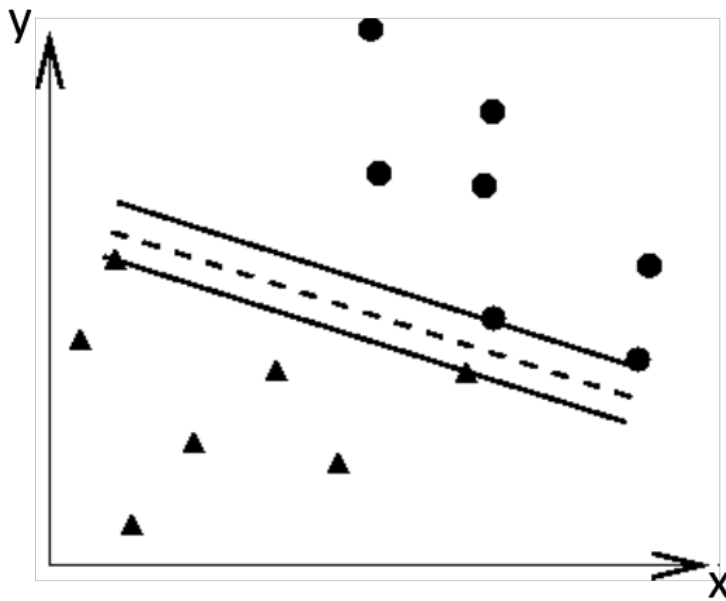
Figure 5.2: An example of how SVM classifies the input data into two classes (quoted from [37])

plex duration, PR interval, QT interval, ST interval, ST segment, RR interval and PR segment as the extracted features and feed these feature to a neural network classifier to detect various heart diseases. There are also some works using Fast Fourier Transform to extract the frequency domain feature [49], or using wavelet transform for feature extraction [92, 53]. Although these works have shown a very good detection accuracy over different heart conditions, the approaches in the existing works can not be used in this research as they are all too complex to be run on mote.

5.2.1 Discussion

For the proposed *AID*, it is designed to work with raw signals in time domain. To improve the detection accuracy of *AID* by using a machine learning classifier, it is also required that the classifier can work with raw signal. When performing classification using raw signals, the neural network approach is believed to be more suitable than SVM approach for the following reasons. The classification using SVM is based on that the SVM can find a hyper-plane and the maximum margins to either classes in the feature space. For the raw signal, each sampling point

can be used as a feature input. In this research, that means there are more than three hundred input features. Subsequently, the hyper-plane will be constructed in more than three hundred dimension. As a result, it is extremely hard to construct the hyper-plane and subsequently hard to be used for classification. On the other hand, the neural network simulates how a human brain works when facing a raw signal. It is capable to find the relationship between sampling points although the relationship known by the neural network can not be understand by human. Furthermore, although neural network is often criticised by its complicated and slow training process, the calculations of the neural network when performing classification is actually very lightweight (only addition and multiplication is involved). As the training process can be run off mote in this research, it will not be a problem preventing the neural network from being useful. Subsequently, the neural network approach will be used in this research.

5.3   AID with Feedforward Neural Network

In this section, the methodology of using FeedForward Neural Network (FFN) to perform the classification in *AID* will be presented. It will start with introducing how the *AID* will be modified to adopt the FFN classification approach. Then, as the neural network need to be trained before it can be used, how the training data is generated in this research will be introduced in Section 5.3.2. The evaluation of the *AID* with FFN classifier will then be presented in Section 5.4. To be easier and clearer to explain the methodology, the modified *AID* using FFN classifier will be named as *FFNAID* in the rest of this section.

5.3.1   Modification towards AID

The proposed *AID* follows a 3-stage detection approach. The purpose of each stage has been introduced in Section 3.3.1. When FFN is used, the *FFNAID* will need to follow a 4-stage approach, which also aim at achieving the phase invariant, the size invariant, and the amplitude invariant. The name and purpose of each stage is as below:

- **Stage 1 - R-Peak Detection:** To align the signal using the R-Peak so that

phase invariant is achieved, and to detect the *Blank Signal* anomaly

- **Stage 2 - Size Unification:** Using the RR-Interval prediction algorithm to modify the signal's length so that the size invariant is achieved

- **Stage 3 - Amplitude Normalisation:** Normalise the signal's amplitude so that the amplitude invariant can be achieved

- **Stage 4 - Signal Classification using FFN:** Classify the signal using the FFN

Each of these stages will be discussed in detail in the rest of this section.

Stage 1 - R-Peak Detection

For the *FFNAID*, the detection is still perform in time domain. Subsequently, it is still important to align the signal, which is waiting for being detected, with the signals used to the training the FFN so that the phase invariant can be achieved. Meanwhile, the feature of *Blank Signal* does not change so that same approach can be used to detect it. Therefore, the purpose of Stage 1 in *FFNAID* remains the same as the *AID*. The *FFNAID* also uses the R-Peak to align signals. The method used to detect the R-Peak is also the same as in the *AID*.

Stage 2 - Size Unification

The purpose of Stage 2 in *FFNAID* is still to achieve the size invariant. However, in contrast to the *AID*, this purpose is not achieved through predicting the **Predicted RR-Interval** based on the **Correct RR-Interval**. Instead, it unifies the size of all RR-Intervals to a pre-set value which is the number of input neurons in the FFN. For a FFN, the number of input neurons is fixed when the network is designed. As a consequence, the size of the input signals will always to be the same for both the training signals or the signals waiting for being detected.

To unify the size of the RR-Interval, the idea of RR-Interval prediction algorithm from the *AID* is borrowed. In the RR-Interval prediction algorithm, the size difference between the **New RR-Interval** and the **Correct RR-Interval** is measured.

Then, the algorithm will extend or shrink the size of the non-cardio-activity duration according to the measured size difference so that the **Predicted RR-Interval** is obtained. In *FFNAID*, instead of measuring the size difference between the **New RR-Interval** and the **Correct RR-Interval**, the algorithm measure the size difference between each RR-Interval and the number of input neurons in the FFN. Then, it will will extend or shrink the size of the non-cardio-activity duration of that specific RR-Interval according to the measured size difference so that all the RR-Intervals' size are unified. In this case, the name of the algorithm is changed to size unification algorithm instead of RR-Interval prediction algorithm.

Stage 3 - Amplitude Normalisation

The purpose of this stage is to achieve the amplitude invariant. In the *AID*, the amplitude invariant is achieved by the Pearson's Correlation Coefficient, which can normalise the difference in amplitude between two signals. However, the Pearson's Correlation Coefficient and threshold approach is replaced by the FFN which does not have the ability to tolerate the difference in amplitude unless it has been trained with different amplitude. In order to minimise the number of training data, this stage is added ito normalise the amplitude difference between any two RR-Intervals.

To normalise the amplitude, the algorithm measure the minimum value in the signal. Then, the signal is shifted up or down so that the minimum value is equal to zero. After this, the maximum value of the shifted signal is measured. Each data point is then divided by the measured maximum number. Subsequently, the whole signal will be normalised with the amplitude between 0 and 1.

Stage 4 - Signal Classification using FFN

In the final stage, the aligned and normalised signal is fed to the neural network's input neuron. The design of the neural network is as below. In this research, the number of input neurons is designed to be equal to the number of data point of a one second ECG signal. For example, the data in MIT-BIH database is sampled at $360Hz$. In this case, the FFN is set to have 360 input neurons as there is 360 data

samples for one second of signal. The reason for one second is that a normal period of ECG signal is around one second (HR is 60 bpm). In this case, each of the ECG signal components will not overlap with each other. If the FFN is designed to have less input neurons, the advantage will be the decrease number of calculations that is needed. Subsequently, the FFN is more lightweight in computation. However, the disadvantage is that the ECG components may be overlapped with each other so that it may become harder for the FFN to classify. On the other hand, if the FFN is designed with too many input neurons, it will take more time and resource to do the computation and store the FFN parameters. Furthermore, the extra input neurons are only used for considering whether the data samples in the non-cardio-activity duration are normal or abnormal. Subsequently, it is meaningless to add more input neurons.

The number of output neurons depends on the number of possible classification results and the way each classification results is represented. In this research, there are three type of anomaly. As the *Blank Signal* anomaly is detected in the Stage 1, In this stage, the FFN will need to classify the input signals into three categories: signal without usage-related anomaly, signal contains *Inverted Signal* anomaly, and signal contains *Noisy Signal* anomaly. To represent the three categories, there are also three possible approaches that are available. First approach is to assign each category a unique label. For example, the signal containing no usage-related anomaly is assigned with label 0, *Inverted Signal* anomaly is assigned with label 1, and *Noisy Signal* anomaly is assigned with label 2. In this case, only one output neuron is needed to represent all three categories. Second approach is binary coding approach. There are three categorises so two binary digits is enough for coding each of the category. For example, the signal containing no usage-related anomaly is coded as 00, *Inverted Signal* anomaly is coded as 01, and *Noisy Signal* anomaly is coded as 10. In this case, each binary requires one output neuron so that two output neurons are needed. The third approach is to use three output neurons and each output neuron represent the positive or negative of the detection of each category. For example, output neuron one represents

the signal containing no usage-related anomaly; output neuron two represents *Inverted Signal* anomaly; and output neuron three represents *Noisy Signal* anomaly. If output neuron one is '1', it means the FFN classifies the input signal into the first category, and vice versa. In this research, the third approach is used. There are three output neurons in the designed FFN. Each output neuron represents the signal containing no usage-related anomaly, *Inverted Signal* anomaly, and *Noisy Signal* anomaly respectively. As each category is represented by one neuron, there is a possibility that the output from the FFN shows two or more categories the input signal belongs to, or the input signal belongs to neither of the three categories. This feature is useful for collecting the training data as the two cases where one input signal belongs to two categories or does not belong to any category means the FFN is confused by that input signal. Furthermore, the output neurons are connected to the last layer of hidden neurons, which is significantly less than the number of input neurons. As a result, having more output neurons than previous two approaches will not create a huge overhead in computation.

For the hidden layer, there is no definite rule that can be followed to decide how many hidden layer is needed and the number of neurons for each hidden layer. The main approach is to set up the FFN with different hidden layer configurations and compares the results from those configurations. Thus, the design of hidden layer will not be covered here. Instead, several different hidden layer configurations will be evaluated in Section 5.4.

5.3.2 Generating Training Data At Run-time

The neural network is a supervised learning approach. Ideally, the supervised learning approach requires the training data to contain all possible patterns that need to be classified. However, as introduced in Chapter 3, it is hard to comprehensively cover all possible patterns of the usage-related anomaly ahead of deployment. Subsequently, an alternative solution need to be used here to generating the training data during the run-time so that the neural network can be retrained using the generated training data, and the classification accuracy can then become more and more accurate during the run-time.

In Section 5.1.1, it has been discussed that if the *AID* need to tolerate more correct templates so that the detection accuracy can be further improved. Due to the on-mote resource limitation especially for the memory, such approach is not feasible. However, for the purpose of generating training data, it can be run off-mote. Subsequently, the resource limitation is not a primary consideration. Thus, the method of incorporating more correct signals' template to *AID* is used here for the purpose of generating training data.

For each confused RR-Interval or the RR-Interval being confirmed with false detection by the *FFNAID*, the training data generator will compare the RR-Interval to all the correct RR-Interval templates in the training data generator's database using the Pearson's Correlation Coefficient. How the database of correct RR-Interval templates is set is outside the scope of this research. Thus, it will not be discussed here. The highest coefficient will then be used to decide which category the RR-Interval belongs to using threshold approach. For example, a confused RR-Interval is reported by the *FFNAID* and it is uploaded to the sink, where the training data generator is running. The generator will calculate the Pearson's Correlation Coefficient between the confused RR-Interval to all the correct RR-Intervals in the correct template database. Assume there are 20 templates in the database and the highest correlation coefficient is obtained with template$\#8$. Then, this correlation coefficient between the confused RR-Interval and the correct template$\#8$ is used to decide which category the confused RR-Interval belongs to. As with *AID*, a threshold, T_C, is used here. If the correlation coefficient is greater than T_C, it means the confused RR-Interval contains no usage-related anomaly. If the correlation coefficient is smaller than $-T_C$, it means the confused RR-Interval contains *Inverted Signal* anomaly. Any value in between T_C and $-T_C$ means the confused RR-Interval contains *Noisy Signal* anomaly.

It has been argued in Section 3.3.1 that a semi-supervised anomaly detection approach will be more suitable for detecting the usage-related anomalies for an ECG-based HC-BSNs. Such argument is still believed to be true here. By following above training data generating approach, the proposed *FFNAID* still remains a

semi-supervised anomaly detection approach although the training of neural network follows the supervised learning approach.

5.4 Evaluation

In this section, the evaluation of the proposed *FFNAID* will be presented. The evaluation is carried out in Matlab due to the lack of suitable hardware equipment. This section will start with introducing the evaluation setup for evaluating each objectives in this research.

5.4.1 Evaluation Metrics

The main focus of this evaluation is the detection accuracy, which is the same when evaluating the *AID*. Subsequently, the evaluation metrics used in this evaluation are still the same. They are true-positive rate (abbreviated as TP), true-negative rate (abbreviated as TN), false-positive rate (abbreviated as FP), false-negative rate (abbreviated as FN), and recognition rate (RR). The detail definition of each of them has been covered in Section 3.4.1. It will not be repeated here.

5.4.2 Evaluation Setup

The main technique used to evaluate the proposed *FFNAID* will still be the same as evaluating the *AID*. It uses the fault injection approach which inject the usage-related anomaly into the source signal so that the evaluation environment can be well controlled and the evaluation metrics can be obtained more accurately. The anomaly database and the source database also remain the same as the evaluation of *AID*. Beside these general evaluation setup, Section 5.1.3 has specified three objectives that need to be achieved and demonstrated through the evaluation. The evaluation for each objective is different to others as they are for different purpose. Therefore, the evaluation setup for each objective will be introduced separately.

Generating training data

In the methodology of generating training data introduced in Section 5.3.2, it requires a correct signal database to be built first. To built such database in this

research, it is assumed that all the RR-Intervals in the source signal database are free from anomaly. The RR-Intervals from each signal in the source signal database will be randomly selected and added to the database and form the correct signal database. Then, RR-Intervals with anomalies being injected will be added to the training data using the methodology introduced in Section 5.3.2. To balance the training data between different classes, for each one correct RR-Interval in the training data set, one RR-Interval contains *Inverted Signal* anomaly and one RR-Interval contains *Noisy Signal* anomaly are also added to the training data set. Overall, the ratio of different classes in the training data set is $1 : 1 : 1$ for RR-Interval which is free from anomaly, which contains *Inverted Signal* anomaly, and which contains *Noisy Signal* anomaly respectively. The total size of the training data set is depending on the setup of each evaluation.

Evaluate the effects of different configuration

For *FFNAID*, the detection accuracy is depending on two configurations. The first configuration is the neural network size. The second configuration is the size of training data. To evaluate each configuration, the other one will have to remain unchanged so that the results are comparable. As there is no definite rule to decide each of these two configurations, therefore, this evaluation will start with evaluate the effect of the neural network size. That means the size of training data will need to be arbitrarily set first. Then, with the evidence of which neural network size will work best for this application, the effect of different size of training data will be evaluated.

To evaluate the effect in terms of detection accuracy when the neural network is configured with different network size, the size of training data is set to 600 templates (200 templates for each classes) for this evaluation. These training data will be used to train the neural network with different size configuration. During the evaluation, anomalies will be randomly injected into the test signal by the fault injection mechanism with the type of the injected anomaly, location of the injected anomaly, and the length of the injected anomaly being recorded. The same test

Page 154

signal will then be fed into the *FFNAID* with different neural network size configuration and the results are recorded. The evaluation metrics are calculated based on the results from the *FFNAID* and the recorded information of those injected anomalies.

To evaluate the effect in terms of detection accuracy when the neural network is trained with different size of training data, the size of the neural network will need to be fixed. To carry out the evaluation, the network size, which has been shown to have the best performance in previous evaluation, will be used. In order to simulate the real application scenario in this evaluation where the methodology collect the information during run-time and using the information to improve the detection accuracy, the evaluation starts with an untrained neural network. The signals with the anomaly injected is fed to the *FFNAID*. The detection result of each RR-Interval is compared with the recorded information of the injected anomaly. For each false result, that specific RR-Interval is fed to the run-time training data generating methodology. Subsequently, the training data database will be gradually built. The neural network will not be retrained until the anomaly detection has been performed against entire signal so that the evaluation metrics can be calculated for that specific number of training data. Then, to evaluate the next configuration of the number of training data, that specific number of training data will be randomly selected from the training data database. These selected data will be used to train the neural network which is then used in the *FFNAID*. The above process is repeated for every configuration of the number of training data until a clear trend can be seen from the results.

Evaluate the detection accuracy improvement

To demonstrate the detection accuracy improvement achieved by *FFNAID* over *AID*, the finding from previous evaluation will be used. More specific, the neural network will be configured with the size that has been demonstrated to have the best performance in previous evaluation. The neural network will then be trained with the number of training data which has also been demonstrated in the previous

evaluation to be with the highest accuracy. After it is trained, the test signals which have been used to evaluate the *AID* in Section 3.4 will be used here to evaluate the *FFNAID*. By doing this, it aims at ruling out the possible detection accuracy difference caused by the different test signals. As the anomaly information for the test signals is known, the evaluation metrics can be obtained by comparing the detection results from *FFNAID* with the known anomaly information in the test signals. Subsequently, the detection accuracy achieved by the *FFNAID* is comparable to the detection accuracy achieved by the *AID*.

Evaluate the necessity of data sharing

The necessity of data sharing in this research is solely referred to whether sharing the training data obtained from subject A can have any benefit in terms of detection accuracy to the anomaly detection carried out on subject B. An real world example scenario is given as follow. It is assumed that User-A has been using the ECG-Based HC-BSNs for a long time whereas User-B is a new user. As User-A has been using the device for a long time, there has been enough training data being obtained and the *FFNAID* on mote can perform accurate usage-related anomaly detection. If the data sharing is beneficial, it is expected that the *FFNAID* for User-A can be used for User-B with similar detection accuracy, or the training data from User-A can be used to train the *FFNAID* for User-B. Otherwise, it is not necessary to share data between User-A and User-B.

Following such definition of the necessity of data sharing, the evaluation is set to carried out in the following way. The neural network is configured using the size which has shown to have the best performance in previous evaluation, and it is trained with the number of training data which has also been shown to be accurate. Then, the test signal, which is using the same source signal as the training data, will be fed to the *FFNAID*. The evaluation metrics is calculated based on the results. The false result will then be added to the training data and the neural network will be retrained. Then, the test signal, which the anomaly is injected into another source signal, will be fed to the *FFNAID*. The results from both tests will

Page 156

Configuration	Comment
FFN10	One hidden layer with 10 neurons
FFN20	One hidden layer with 20 neurons
FFN(10-2)	Two hidden layer with 10 neurons in first hidden layer and 2 neurons in the second hidden layer
FFN(10-5)	Two hidden layer with 10 neurons in first hidden layer and 5 neurons in the second hidden layer
FFN(10-10)	Two hidden layer with 10 neurons in first hidden layer and 10 neurons in the second hidden layer

Table 5.1: The evaluated neural network size configuration

be compared. To show the outcome is applied to most source signals, the above test will be repeatedly carried out against all the source signals.

5.4.3 Evaluation Result

In this section, the evaluation results will be presented following the setup of each evaluation. The sequence that the results are presented are the same as the sequence of each evaluation was carried out.

Evaluation result for network size configuration

The first evaluation that has been carried out is to demonstrate the effect in terms of the detection accuracy with different neural network size configuration. The following neural network size configurations and their representation in the plotted results are listed in Table 5.1.

The evaluation results for different neural network size configurations have been plotted in Figure 5.3, Figure 5.4, Figure 5.5, and Figure 5.6 in the form of a box-plot. First of all, for the true negative rate (shown in Figure 5.3), it can be seen that for those configurations, which only have one hidden layer, perform better than those configurations which have two hidden layers. For the two candidate configurations that have one hidden layer, it can be seen that the detection accuracy is more unified over different source signal if the neural network has 10 hidden neurons. On the other hand, there is much more statistical outliers for the true negative rate achieved over different test signals when there are 20 hidden neurons in the neural

	p-Value
True Negative Rate	0.4241
True Positive Rate	0.5592
False Negative Rate	0.3292
False Positive rate	0.5533

Table 5.2: Statistic significance test between FFN10 and FFN20

network. For the true positive rate (shown in Figure 5.4), it can also be seen that one hidden layer configurations are better than two hidden layer configurations. Among the two one hidden layer configurations, it can be seen from the boxplot that FFN20 has less variance between different test signals but there are more statistic outliers than FFN10. For the false negative rate, it can also be seen that the one hidden layer configurations are slight better than the two hidden layers configurations. The FFN20 may have smaller variance than FFN10 but if taking account the statistic outliers, these two configurations achieve similar false negative rate. In terms of the false positive rate (shown in Figure 5.6), it can be seen that there is no significant difference between each configuration except FFN(10-5). However, if considering the statistic outliers, it can be clearly seen that FFN10 is better than the other configurations.

Based on the evaluation results for different neural network size configurations, it can be concluded that the one hidden layer configurations is better than two hidden layer configurations. For FFN10 and FFN20, a statistic significance test has been carried out to see whether the difference between those rates achieved by FFN10 and FFN20 are significant. The results has been shown in Table 5.2. It can be seen that the p-Values for those rates are all bigger than 0.05, which means there is no significant difference. Subsequently, the rest of this research will use the FFN10 configuration for the consideration of calculation complexity.

Evaluation result for training data configuration

After the neural network size configurations have been evaluated, the next evaluation looks into the effect of different number of training data. The evaluation has been run with the neural network being trained using 1 to 1000 templates with 50

Figure 5.3: Boxplot of the true negative rate achieved by the *FFNAID* with different size configuration. From the left to right, it shows the achieved rate for FFN10, FFN20, FFN(10-2), FFN(10-5), and FFN(10-10). In general, the FFN10 configuration can achieve better true negative rate with less variance compared to other configurations.

Figure 5.4: Boxplot of the true positive rate achieved by the *FFNAID* with different size configuration. From the left to right, it shows the achieved rate for FFN10, FFN20, FFN(10-2), FFN(10-5), and FFN(10-10). In general, the one hidden layer configurations can achieve better true positive rate compared to two hidden layers configurations. However, the two one hidden layer configurations achieve similar true positive rate but FFN20 configuration has less variance but more statistical outliers.

Figure 5.5: Boxplot of the false negative rate achieved by the *FFNAID* with different size configuration. From the left to right, it shows the achieved rate for FFN10, FFN20, FFN(10-2), FFN(10-5), and FFN(10-10). In general, the one hidden layer configurations can achieve better false negative rate with less variance compared to two hidden layers configurations. There is not much difference between the two one hidden layer configurations.
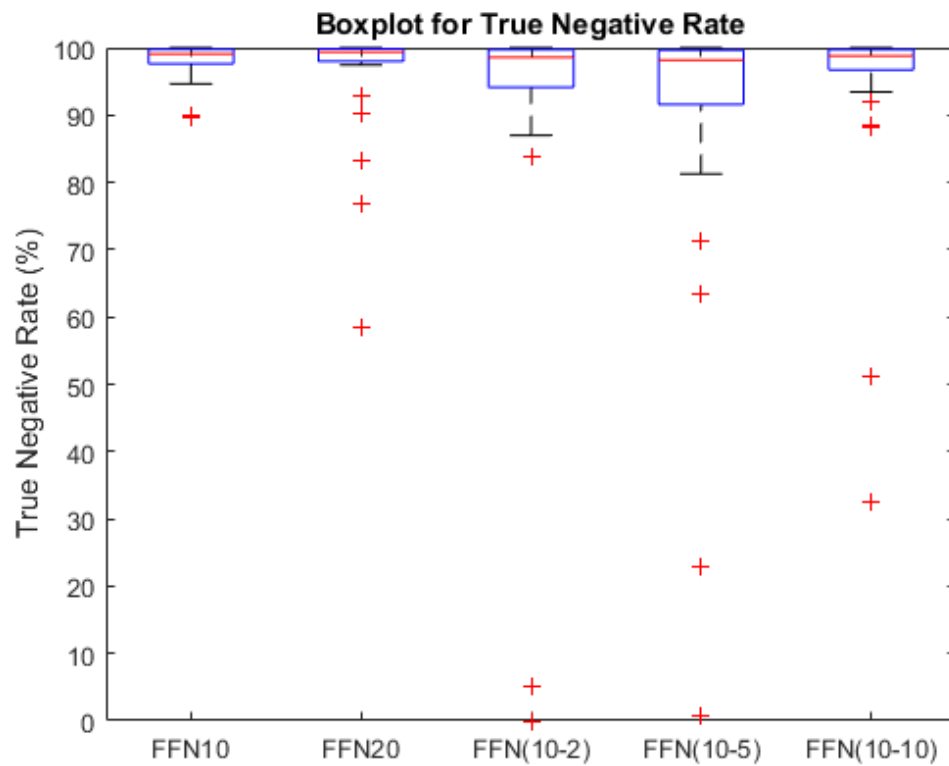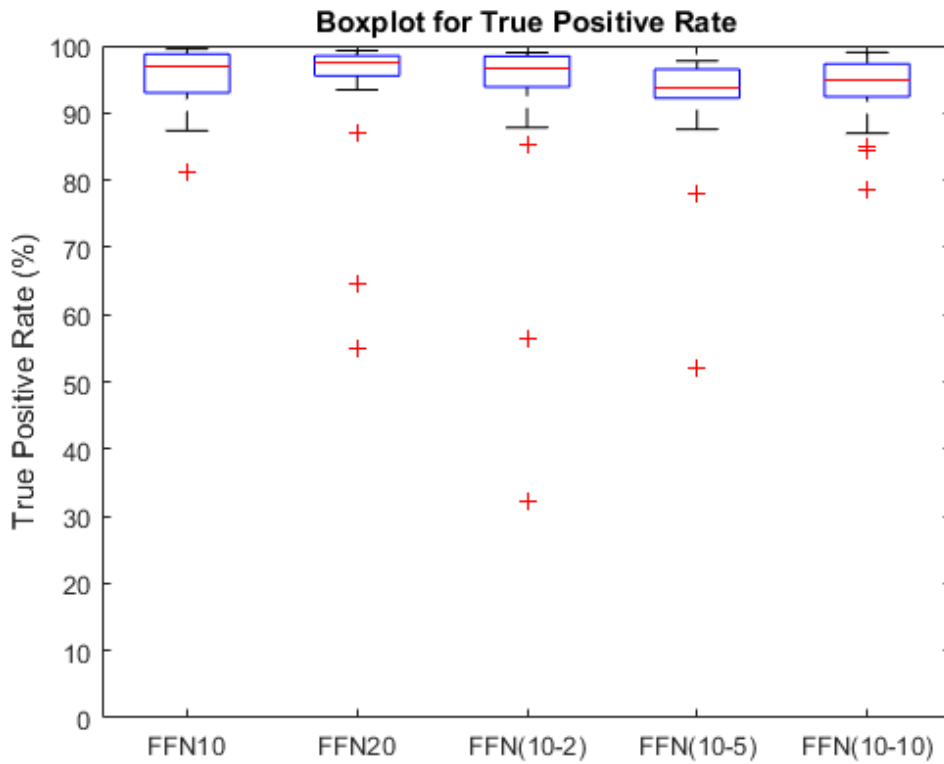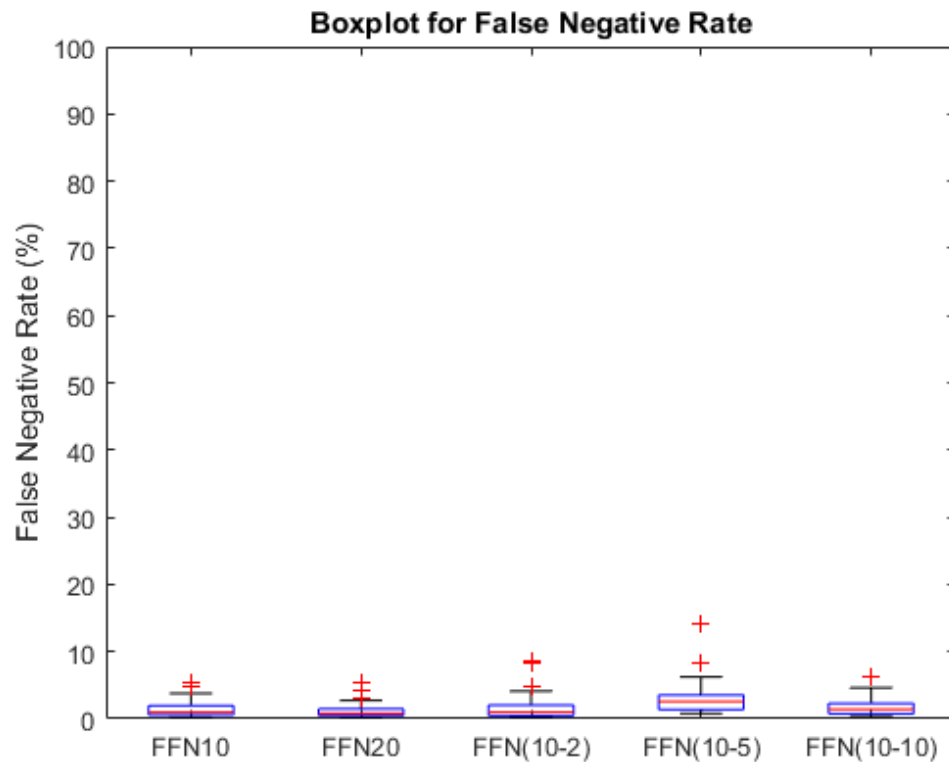
Boxplot for False Positive Rate
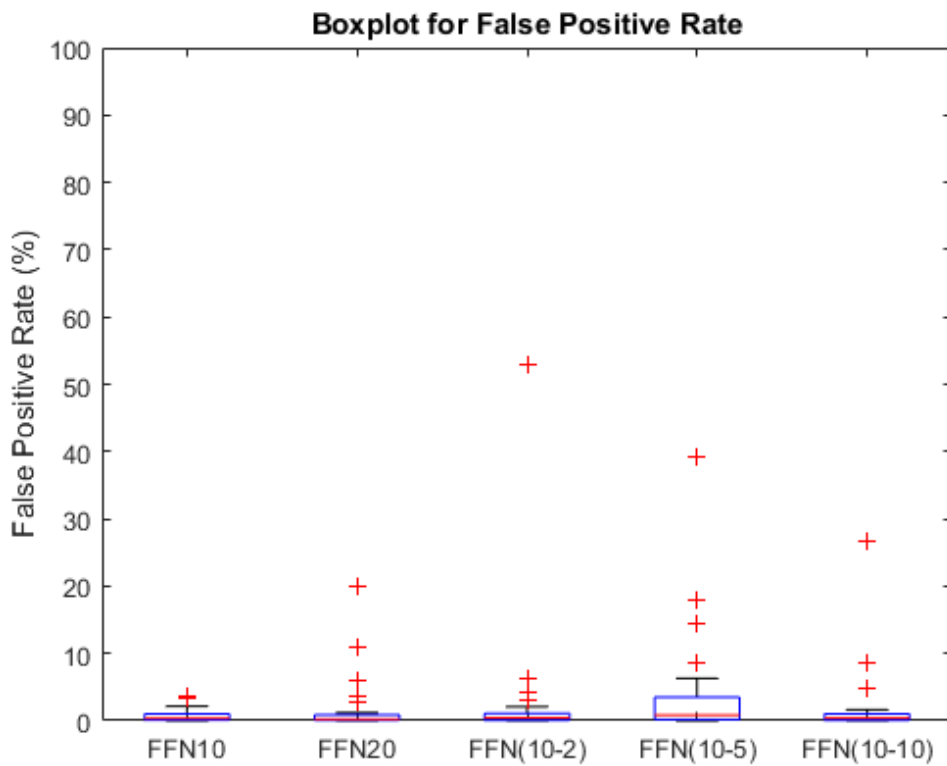
Figure 5.6: Boxplot of the false positive rate achieved by the *FFNAID* with different size configuration. From the left to right, it shows the achieved rate for FFN10, FFN20, FFN(10-2), FFN(10-5), and FFN(10-10). In general, the FFN10 configuration can achieve better false positive rate with less variance compared to other configurations.
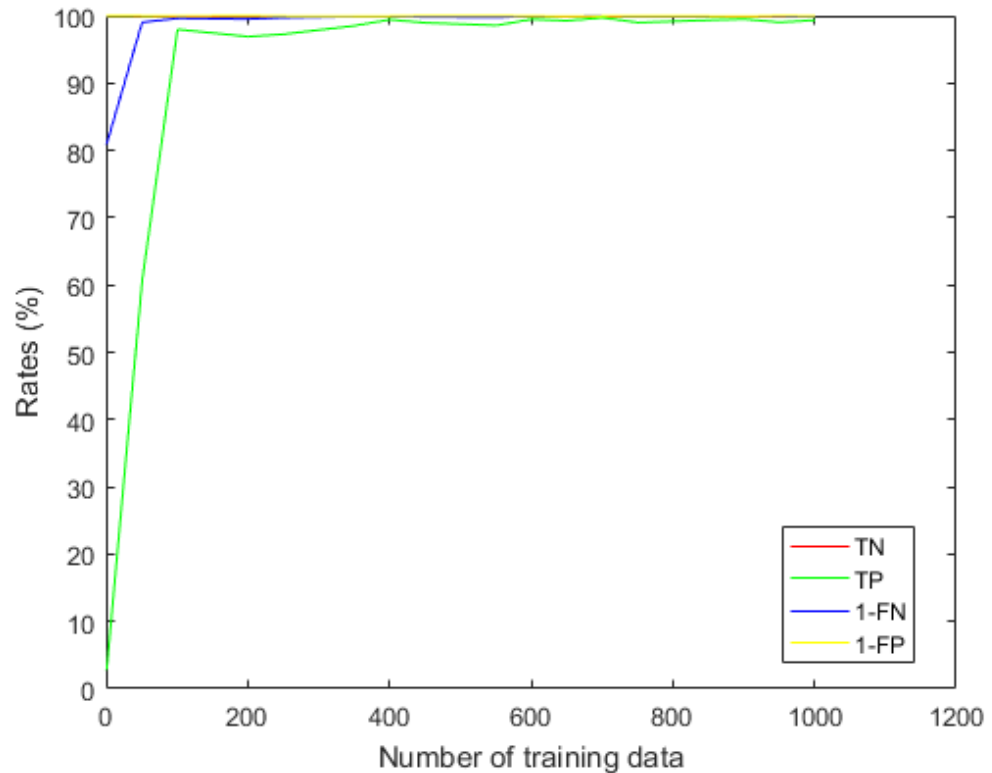
Figure 5.7: Evaluation result for the effect of different number of training data.

increment. The result is shown in Figure 5.7 and Figure 5.8. Figure 5.7 shows the overall result from this evaluation. It can be seen that if the number of training data is lower than 100, the detection accuracy is significantly poor. After the number of training data reaches 100, detection accuracy improvement can still be seen but the improvement is less dramatical than before. Figure 5.8 shows a zoom-in version of the result from this evaluation which range from 150 training data to 1000 training data. It can be seen from this plot that with the number of training data increase, the detection accuracy is improve gradually. The maximum accuracy is achieved when the number of training data reaches 700. There is some slight variation for each rate after the number of training data reaches 400. It is believed to be caused by the stop training conditions for the neural network which aims at preventing the overfitting.

Figure 5.8: Zoom-in version of the evaluation result for the effect of different number of training data. It can be seen that the detection accuracy is improved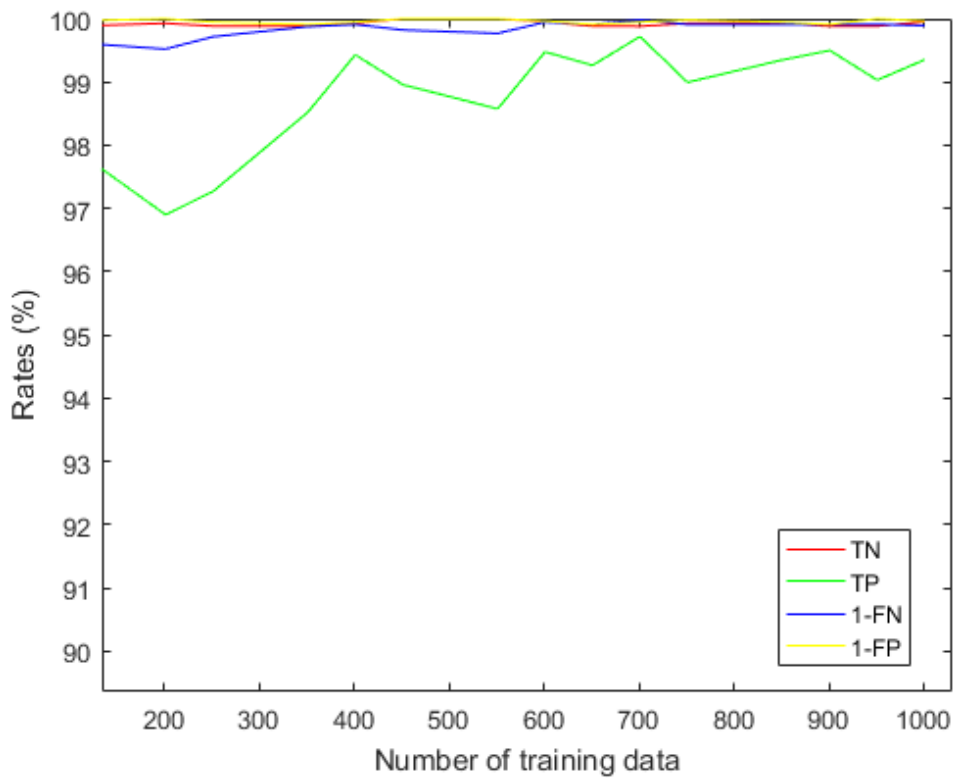 with the number of training data increase. However, once the number of training data reaches a certain point (in this case, 400 training data), the more training data will not contribute to the improvement of detection accuracy significantly.

Evaluation result for detection accuracy improvement

The third evaluation is to demonstrate the improvement in terms of detection accuracy achieved by *FFNAID* compared to *AID*. Based on the finding from previous evaluations, this evaluation is carried out with FFN10 configuration and for each test signal, and the neural network will be trained with 400 training data. During the research, *Inverted Signal* anomaly and *Noisy Signal* anomaly with noise strength of $-10dB$, $-6dB$, and $0dB$ have been evaluated. The results from this evaluation have been shown in Table 5.3.

When the *Inverted Signal* anomaly is injected, the *FFNAID* can improve the true positive rate, true negative rate, false positive rate, and false negative rate by 1.36%, 2.31%, 0.64%, and 0.26% respectively compared to *AID*. The p-Values from statistic significance test for each rate are below 0.05. That means the improvement achieved by *FFNAID* is significant.

When the *Noisy Signal* anomaly with noise strength of $-10dB$ is injected, *FFNAID* can achieved 0.35%, 5.04%, 2.58%, and 0.59% improvement over *AID* for true positive rate, true negative rate, false positive rate, and false negative rate respectively. However, the statistic significance test shows that the improvement of true negative rate is not statistic significant as the p-Value is 0.217 which is higher than 0.05. For other three rate, the statistic significance test has shown that the improvement is significant.

When the *Noisy Signal* anomaly with noise strength of $-6dB$ is injected, it can be seen from the results that the *FFNAID* can achieve significant improvement for true positive rate, true negative rate, and false positive rate, which are 9.15%, 5.04%, and 2.19% respectively. However, the result has also shown a significant worse false negative rate. Compared to the *AID*'s 0.9% false negative rate, the *FFNAID* achieved 2.45% which is 1.5% worse than *AID*.

Similar results can also be observed for the *Noisy Signal* anomaly with $0dB$ noise strength. The *FFNAID* can achieve 24.23% true positive rate improvement com-

pared to *AID* while the true negative rate and false positive rate are improved by 6.47% and 2.52% respectively. These improvements are all significant as the p-Values are all less than 0.05. However, the false negative rate is worsen by 4.47% compared to *AID* and the statistic significance test has shown that it is significant.

To investigate what caused the false negative rate become worse compared to *AID*, a further analysis has been carried out. The analysis begins with examining the rate achieved by each test signal. As the test signals are obtained through injecting anomaly into the source signals, the second step has examined the signal quality of the source signals. Similar to the situation when evaluating *AID*, the source signals may also contain certain amount of noise. However, as it has recovered after a certain time and it does not affect the rest signal to be useful to the cardiologist, these signals are remained in the database. When organising the training data, it assumed that all RR-Intervals from the source signals are free from anomaly as annotating each RR-Interval for its quality in this research is not feasible due to the amount of labour required. Subsequently, if the source signal has already had certain amount of noise, it will train the neural network to classify it as free from anomaly. On the other hand, such worsen results in false negative rate can only be seen when the injected *Noisy Signal* is weak. Through plotting both the test RR-Interval and the source signal that has already contained noise, it can be observed that they are similar in time domain. As a result, the false negative result is obtained. To further proof above analysis, the result which was obtained using the source signal that is fully clean has also been examined and compared. In this case, the MIT-BIH$\#100$, which is totally free from any anomaly, and MIT-BIH$\#122$, which has already contained certain amount of noise, are used as the example. The false negative rate achieved, when *Noisy Signal* anomaly with strength of $0dB$ is injected into MIT-BIH$\#100$, is 0.4%. On the other hand, the false negative rate achieved, when *Noisy Signal* anomaly with strength of $0dB$ is injected into MIT-BIH$\#122$, is 32.6%. Subsequently, it is believed that the above analysis of why the false negative rate is worsen compared to *AID* is true. To overcome such issue, one solution may be able to adopt is to use a better signal quality check algorithm

Injected Anomaly		TP	TN	FP	FN
	FFNAID	99.88%	99.9%	0.02%	0%
Inverted	AID	98.52%	97.59%	0.66%	0.26%
Signal	Improvement	1.36%	2.31%	0.64%	0.26%
	p-Value	0.04	0.001	$9.5e-05$	0.04
Noisy	FFNAID	99.29%	100%	0%	0.18%
Signal	AID	98.94%	94.96%	2.58%	0.77%
$-10dB$	Improvement	0.35%	5.04%	2.58%	0.59%
	p-Value	0.2	$3e-08$	0.001	$4.7e-06$
Noisy	FFNAID	93.36%	99.88%	0.04%	2.45%
Signal	AID	84.21%	94.84%	2.59%	0.9%
$-6dB$	Improvement	9.15%	5.04%	2.19%	-1.5%
	p-Value	$4.7e-04$	$9.6e-08$	0.003	$1.1e-07$
Noisy	FFNAID	76.15%	99.43%	0.12%	7.3%
Signal	AID	51.92%	92.96%	2.64%	2.83%
$0dB$	Improvement	24.23%	6.47%	2.52%	-4.47%
	p-Value	0.004	$7.7e-06$	0.04	$1.2e-05$

Table 5.3: The evaluation results shows detection accuracy achieved by *FFNAID*. The results are compared with the detection accuracy achieved by *AID*. Overall, the *FFNAID* can achieve better than *AID*. However, when the *Noisy Signal* anomaly is presented and the noise strength is weak, the *FFNAID* will have higher false negative rate.

off-mote so that more accurate data labels can be assigned when organising the training data. Subsequently, the on-mote neural network can be trained to a higher accuracy so that the FN rate can be further improved. However, the research of a better off-mote signal quality check algorithm is out of the scope of this research. Thus, such solution is proposed here as an example and for the future work.

Evaluation result for data sharing

The results of sharing data between different subjects have been shown in Figure 5.9. In each figure, the red '∗' means the rate achieved by one specific subject. The horizontal axle represents the number of training data. As after the *FFNAID* finishes the detection for the test signal from one subject, the false result will be added to the training data with correct annotation and the neural network will be retrained before the *FFNAID* being used for the signal from next subject. Subsequently, when completing the detection for each subject, the number of training

data will increase.

From these plots it can be seen that there are quite significant variation for each rate when the detection for current subject uses the *FFNAID* trained by previous subjects' training data. When the number of training data reaches more than 4500, it seems the variation becomes less but with one exception. Through checking the signals' appearance, it can be observed that for those close detection results, the source signals' appearance are quite similar. That means the close results achieved by two neighbour are not benefit from the training data sharing. Instead, whether two detections can achieve similar results using the *AIDFFN* trained with same training data is still heavily depend on the similarity between the signals themselves. From this perspective, it can be concluded that there is no significant benefit to share data between different subject.

On the other hand, if comparing the result achieved using the *FFNAID*, which is trained using others' training data, to the result achieved using the untrained *FF-NAID* (the result is not plotted but the detection accuracy is extremely poor), it can be said that the *FFNAID* trained using others' training data is performed much better than the untrained *FFNAID*. However, it is very rare that no training data can be obtained from a subject to train the *FFNAID*. If there is indeed no training data available, one realistic example will be to use the *AID* to perform anomaly detection and collect training data. The data is then used to train the *FFNAID*. Until a certain point when the two algorithm can achieve similar detection accuracy or the *FFNAID* can achieve better accuracy, the system can switch on the *FFNAID* and switch off the *AID* for the future detection. Based on the evaluation results and above analysis, it can be concluded that there is no significant benefit to share data between different monitoring subject in terms of improving the detection accuracy.

Discussion of the overhead

Although this research mainly aims at demonstrating the detection accuracy improvements over the *AID*, it is still necessary to discuss the possible overhead if
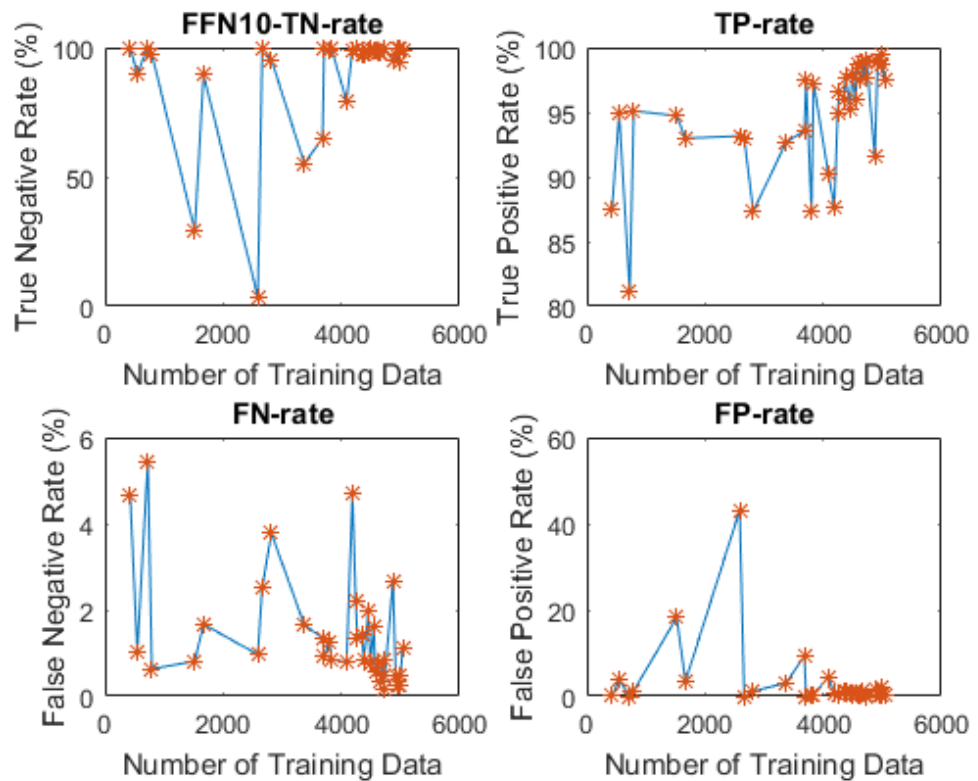
Figure 5.9: The evaluation results of the necessity of sharing data between different subjects. Top-left shows the true negative rate. Top-right shows the true positive rate. Bottom-left shows the false negative rate. Bottom-right shows the false positive rate. The red '∗' in each figure means the rate achieved by one specific subject. From these plots it can be seen that there are quite significant variation for each rate when the detection for current subject uses the *FFNAID* trained by previous subjects' training data.

such approach is adopted. In this section, the overhead will be discussed by theoretical analysis. As introduced in Section 5.1.2, the communication between the mote and the off-mote processing resources (e.g remote servers) can take place when the mote is being charged. Therefore, the communication overhead can be ignored as no on-board energy is consumed in this case. Subsequently, the computation overhead and memory will be mainly discussed here. These overheads will be discussed using the configurations and parameters which have been shown to have best detection accuracy in previous experiments.

For the memory overhead, assuming the signal is sampled at $256Hz$, which meets the minimum ECG signal quality requirement, there will be 256 input neurons. With the neural network configuration of one hidden layer with 10 hidden neurons and three output neurons, it requires 2600 weights to be stored. As it has been previously shown that the feed forward neural network can be implemented on a $8\text{-}bit$ MCU [82], it is assumed that each weight requires $8\text{-}bit$ to store. Thus, the ROM overhead will be 2600 Bytes. For the run-time RAM overhead, the *FFNAID* will only need to store one RR-Interval which is normalised to one second of length (equals to 256 samples if the sampling frequency is $256Hz$). As each sample is with $12\text{-}bit$ sampling resolution, it will generate 384 bytes of RAM overhead. For the computational overhead, each hidden neuron will requires 256 multiplications and 257 additions. Therefore, the hidden layer will requires 2560 multiplications and 2570 additions in total. Then, each of the three output neurons requires 10 multiplications and 11 additions. As a result, the total number of multiplications required is 2590 and the total number of additions required is 2603.

5.5 Summary

This chapter aims at investigating the benefits in terms of detection accuracy if using off-mote collaboration, which is the third research question in this thesis. The research question is quoted below:

> "Can the off-mote collaboration offer any benefits in terms of the detection accuracy?"

On the way to address this research question, this chapter has looked into the limitation of existing *AID* when it reports false results. Based on the limitation, improvement proposal has been drafted accordingly. More specifically, the improvement focus on using machine learning classifier in existing *AID* so that more templates can be learnt and better detection accuracy can be expected. Following the proposal, we have proposed a new methodology based on existing *AID*. We named the new methodology as *FFNAID*. The *FFNAID* has been evaluated using the same method as the *AID*. The evaluation results have first demonstrated how different configurations of the *FFNAID* can affect the detection accuracy. Then, the improvement in terms of detection accuracy over existing *AID* has been demonstrated. On top of these results, this chapter has also looked into the benefit of sharing data between different subject. The result has shown that there is no significant benefit in terms of detection accuracy if the data is shared between different subjects. As this is the final chapter that presents the technical work in this research, the next chapter is going to summarise all the finding in this research and propose the possible future work in this domain.

Chapter 6

CONCLUSION

The proposal of using HC-BSNs to relief the pressure on nowadays' healthcare system is going to be a real contender to the solution for the high pressure on existing healthcare system. However, there are still plenty efforts that are needed to assure the dependability of these systems as an undependable system for healthcare purpose may ultimately lead to the death of the patient. This thesis has highlighted one of the issues which can have significant impact to a system's dependability. We referred this issue to the usage issue, which is defined as whether a system is being used correctly. To investigate the usage issue, this thesis used an ECG-Based HC-BSNs as an example and a systematic approach has been adapted to address the usage issue. It aimed at detecting when the system is being erroneously used through the captured anomalous signals, and identifying what is the root cause to those anomalous signals. By knowing these information, it is believed that the usage issue can be address through prompting the user with the erroneous usage, and assisting the user to rectify the usage issue.

This thesis has presented three methodologies to addressing the usage issue of an ECG-Based HC-BSNs. Two of them aim at detecting the usage-related anomaly and the other one aims at identifying the root cause to the usage-related anomaly. The contributions that have been made in this thesis is summarised as follows:

- **Contribution 1 - On mote run-time usage-related anomaly detection**

  In Chapter 3, the first research question in this thesis was addressed by proposing an on mote run-time usage-related anomaly detection algorithm. Through a number of evaluations using different setups and signals from different subjects, the proposed algorithm has demonstrated that the anomalous signals caused by the erroneous usage of the device can be detected on a mote at run-time.

- **Contribution 2 - Define the relationship between usage issues and on mote signals**

  In Chapter 4, on the way to identify the root cause to the usage-related anomaly, it requires the knowledge of how each of the possible usage issue can affect the signals on a mote platform. However, there is no existing work for this purpose. Subsequently, a qualitative study has been carried out to define the relationship between usage issues and the on mote signals.

- **Contribution 3 - Root cause identification with out human intervention**

  In Chapter 4, the second research question in this thesis was addressed through proposing a root cause identification methodology. Through evaluating the proposed methodology using real data captured from volunteer participants, the results have demonstrated that the root cause to the usage-related can be identified without human intervention. The evaluation of the proposed methodology has also demonstrated that the identification accuracy is heavily depending on the usage-related anomaly detection accuracy, which subsequently motivated our third technical work.

- **Contribution 4 - Improving usage-related anomaly detection accuracy using off-mote collaboration**

  In Chapter 5, the third research question in this thesis was addressed through proposing a new methodology which can learn many templates off-mote so that the detection can be more accurate. The evaluation of the methodology has demonstrated the improvement in detection accuracy over existing

methodology. At the same time, the evaluation has also looked into how the different configuration of the methodology can affect the detection accuracy, and the benefit in terms of detection accuracy if sharing data between different subjects. For the later one, the result has shown that sharing data between different subjects has no significant benefit in term of the accuracy when detecting the usage-related anomaly.

6.1 Limitations

Despite the contributions listed above, there are still some limitation to address the usage issue. Through review the research works, the following limitations have been concluded:

Root cause identification accuracy

Through the evaluation of the proposed *ACLP*, it has shown that the root cause identification is not accurate when certain types of activities are performed by the monitoring subjects. More specifically, those types of activities can heavily exercise the muscle at the chest area but the body movement intensity is not strong (e.g push-up), which can subsequently lead to the ECG signal being corrupted by EMG signal. The methodology proposed in this research has been shown not to be able to cope with these scenario well. Although some further studies of these scenario have been carried out during the research, it still has not been address yet due to the limitation in experiment equipment. In here, we have identify this limitation and the possible future work will be proposed in next section, which aims at addressing this limitation.

12-Lead ECG devices

In this research, the usage issue was investigated using the 3-Lead ECG-Based HC-BSNs. However, as the literature review has presented, there are other types of ECG devices. Among these types of ECG devices, the 12-Lead ECG device is the most widely used in clinical conditions and it can provide the most comprehensive view of the heart. However, due to the equipment limitation, we have not

yet demonstrated the effectiveness of the methodologies against the 12-Lead ECG device. Although it is believed that the proposed methodologies can be adapted to 12-Lead ECG device, further evaluations are still required to demonstrate the evidence of our claim.

Human involvement

The whole research presented in this research is based on the idea of address the usage issue by detecting the usage-related anomaly in the captured signal and identifying the root cause to the usage anomaly so that the user can be prompted and subsequently rectify the usage issue. Therefore, human is still involved in the process of addressing the usage issue. However, as one of the root cause to the usage-related anomaly is the patient being not cooperative, there is still a chance that the usage-related anomaly is correctly detected and its root cause is identified, but the user does not rectify the usage issue. In this case, the methodologies proposed in this research will still not be able to address the usage issue. As a result, we consider that the human involvement will be a limitation of the work presented in this thesis.

6.2 Future Work

Based on the limitations of the work in this thesis, the following future works are proposed:

Improving root cause identification accuracy

As been previously mentioned, the root cause identification accuracy is suffered when the ECG signal is corrupted by EMG signal, and the EMG signal is caused by the activities which heavily exercise the muscle but do not have intensive body movement. Push-up will be a very good example in this case. If the push-up is performed slowly (around 1 push-up per second) and smoothly, it can exercise the chest muscle intensively. However, as it is performed smoothly, the accelerometer can not detect intensive body movement. As a consequence, the proposed *ACLP*

can confuse with the root cause. To address this limitation, two directions can be taken.

First direction will be to recognise a specific range of activities which can currently confuse the *ACLP*. By recognising those activities, the new checks can be added to the checklist so that the identification accuracy can be improved. This direction includes the following two challenges. First challenge here is to define the activities that can lead to the situation where *ACLP* is confused. Second challenge is whether these activities can be accurately detected using signals from accelerometer. For the first challenge, we have explored which muscle's activity can lead to the *ACLP* being confused. However, it is now known which activity can lead to this situation. To improving the identification accuracy by recognising these activities, it is necessary to have a comprehensive list of those activities. For the second challenge, existing works on activity recognition may not yet aim at detecting those possible activities. At least there is no literature can be found to detect push-up. Tapia et al. [83] has studied the activity recognition over different gymnasium activities. In their study, they have listed the push-up as one activities. But, they have not evaluated over this activity and no reason has been given. It is suspected that it is due to the difficulty when recognising this activity so they omitted it during evaluation. There is some application on mobile phone which can recognise the subject is doing push-up. However, they are based on proximity sensor.

Second direction will be to using more signal sources to perform the identification. For example, it can use a separate EMG sensor to record the EMG signals. Features can be extracted from the EMG signal and used to identify the root cause. However, it has been reported that the ECG signal and EMG signal can usually corrupt each other. Subsequently, the challenge for this direction will be to find the most appropriate placement of the EMG sensor(s) so that the situations which can confuse the *ACLP* can be fully covered with the minimum number of sensor needed.

There is also a final research question in this future work but it may be outside

the scope of this research. Such question is to investigate whether those situations, where *ACLP* is confused, will be a problem for the real application of the ECG-Based HC-BSNs. For example, in real life, those people, who have been prescribed with using ECG-Based HC-BSNs, are with certain medical conditions. It is possible that these people will never or not able to perform those activities to confuse the *ACLP*. If this is the case, there is no need to spend time on improving the identification accuracy against these situations.

Demonstrate the effectiveness of the *AID* and *ACLP* to 12-Lead ECG-Based HC-BSNs

The second future work proposed in this thesis is to demonstrate the effectiveness of the *AID* and *ACLP* to 12-Lead ECG-Based HC-BSNs. As it is the widely used ECG recording approach in clinical environment, it is believed that it will be adapted to ECG-Based HC-BSNs in the future. The 12-Lead ECG recording requires a more complicated deployment, which means more usage issues can happen. Subsequently, it is necessary to demonstrate the methodologies' capability to address the usage issues of the 12-Lead ECG-Based HC-BSNs.

Using self-organising sensor node to address the usage issue

The final proposal of the future work is to use self-organising sensor node to address the usage issue. One of the features of Wireless Sensor Network is its ability to form a network based on each mote's location. As a result, each mote's location does not need to be engineered before deployment. By adapting such feature to address the usage issue, the following concept is proposed to address the usage issue. In the concept, two types of motes may needed. First type is reference mote and second type is the sensing mote. The reference mote is designed to be attached to the locations where a mistake is less likely to be made (e.g can be attached to belly button and two nipple although it may not be a good example). For the sensing mote, they are designed to perform the same functions (e.g for ECG-Based HC-BSNs, each mote is design to capture the electrical signal on skin). Then, the mote is deployed to body without specifying which one should

Page 178

be attached to which location. The motes will then find out their location based on the reference motes (such process is known as localisation), and each sensing mote will choose its best sensing signal based on the location. By doing this, it can rule out many misplacement or misconnection problems. As the whole process of deciding which mote to sense which signal does not involve the human operation. Subsequently, it is possible to address the usage issue without human involvement.

The challenge for such proposal is the localisation precision. There are many existing works on the wireless sensor network localisation (e.g Yao et al. [90], and Mao [54]). However, as the wireless sensor network is usually deployed to an outdoor and big area, it is not known whether existing approaches can work when the mote is deployed to a human body and whether the precision can support the choose of the sensing signal. If existing works can not work on human body or the precision is not good enough, how to improve these two factors will be the main research question for this proposal.

Bibliography

[1] *Development Needs Report by NITRD, High-Confidence Medical Devices: Cyber-Physical Systems for 21th Century Health Care A Research and Development Needs Report.* Feb. 2009.

[2] Shimmer sensing: http://www.shimmersensing.com/. Website, Feb. 2014.

[3] 10 types of ECG devices for Heart Rhythm Monitoring. https://www.medicwiz.com/medtech/diagnostics/10-types-of-ecg-devices-for-heart-rhythm-monitoring, Apr. 2016. Accessed on 18 September 2016.

[4] M. Al Ameen, J. Liu, and K. Kwak. Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications. *Journal of Medical Systems*, 36(1):93–101, Mar. 2012.

[5] M. AlGhatrif and J. Lindsay. A brief review: history to understand fundamentals of electrocardiography. *Journal of community hospital internal medicine perspectives*, 2(1), 2012.

[6] U. Anliker, J. A. Ward, P. Lukowicz, G. Troster, F. Dolveck, M. Baer, F. Keita, E. B. Schenker, F. Catarsi, L. Coluccini, A. Belardinelli, D. Shklarski, M. Alon, E. Hirt, R. Schmid, and M. Vuskovic. AMON: A Wearable Multiparameter Medical Monitoring and Alert System. *IEEE Transactions on Information Technology in Biomedicine*, 8(4):415–427, Dec. 2004.

[7] D. Arney, J. M. Goldman, S. F. Whitehead, and I. Lee. Synchronizing an X-ray and Anesthesia Machine Ventilator: A Medical Device Interoperability Case Study. In *International Conference on Biomedical Electronics and Devices, BioDevices*, Jan. 2009.

[8] A. Avižienis, C. Landwehr, J.-C. Laprie, and B. Randell. Basic Concepts and Taxonomy of Dependable and Secure Computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1):11–33, Jan. 2004.

[9] O. Aziz, B. Lo, J. Pansiot, L. Atallah, G.-Z. Yang, and A. Darzi. From computers to ubiquitous computing by 2010: health care. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1881):3805–3811, Oct. 2008.

[10] J. Baek, G. Lee, W. Park, and B.-J. Yun. Accelerometer Signal Processing for User Activity Detection. In M. Negoita, R. Howlett, and L. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3215 of *Lecture Notes in Computer Science*, chapter 82, pages 610–617. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[11] P. Baronti, P. Pillai, V. W. C. Chook, S. Chessa, A. Gotta, and Y. F. Hu. Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards. *Computer Communications*, 30(7):1655–1695, May 2007.

[12] F. Belli and R. Crisan. Towards automation of checklist-based code-reviews. In *Proceedings of the 7th International Symposium on Software Reliability Engineering*, pages 24–33. IEEE Comput. Soc. Press, 1996.

[13] R. R. Bond, D. D. Finlay, C. D. Nugent, G. Moore, and D. Guldenring. A simulation tool for visualizing and studying the effects of electrode misplacement on the 12-lead electrocardiogram. *Journal of Electrocardiology*, 44(4):439–444, July 2011.

[14] G. Bortolan and I. I. Christov. Principal Component Analysis for detection and assessment of T-wave alternans. In *Computers in Cardiology*, Sept. 2008.

[15] A. K. Bourke, J. V. O'Brien, and G. M. Lyons. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait & Posture*, 26(2):194–199, July 2007.

[16] G. E. Burch. History of precordial leads in electrocardiography. *European journal of cardiology*, 8(2):207–236, Sept. 1978.

[17] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Survey*, 41(3):1–58, 2009.

[18] M. Chuah and F. Fu. ECG Anomaly Detection via Time Series Analysis. In *Frontiers of High Performance Computing and Networking ISPA 2007 Workshops*, volume 4743 of *LNCS*, pages 123–135. Springer Berlin Heidelberg, 2007.

[19] Crossbow. *TelOS B Datasheet*, May 2013.

[20] A. Dunkels, B. Grönvall, and T. Voigt. Contiki - A Lightweight and Flexible Operating System for Tiny Networked Sensors. In *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, pages 455–462, Nov. 2004.

[21] J. Eldridge, D. Richley, C. Ross, C. Cox, and C. Breen. Clinical Guidelines by consensus. Recording a standard 12-lead electrocardiogram: An approved methodology by the Society for Cardiological Science & Technology (SCST). Technical report, 2014.

[22] M. L. Fairbairn. *Dependability of Wireless Sensor Networks*. PhD thesis, University of York, Sept. 2014.

[23] W. B. Fye. A history of the origin, evolution, and impact of electrocardiography. *The American journal of cardiology*, 73(13):937–949, May 1994.

[24] M. K. Garg, D. J. Kim, D. S. Turaga, and B. Prabhakaran. Multimodal Analysis of Body Sensor Network Data Streams for Real-time Healthcare. In *Proceedings of the International Conference on Multimedia Information Retrieval*, MIR '10, pages 469–478, New York, NY, USA, 2010. ACM.

[25] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23):e215–e220, June 2000.

[26] J. M. Goldman, R. A. Schrenker, J. L. Jackson, and S. F. Whitehead. Plug-and-Play in the Operating Room of the Future. *Biomedical Instrumentation & Technology*, pages 194–199, May 2005.

[27] J. W. Grier. How to use 1-Lead ECG recorders to obtain 12-Lead resting ECGs and Exercise ECGs. Online at https://www.ndsu.edu/pubweb/˜grier/1to12-lead-ECG-EKG.html, Sept. 2008.

[28] J. Guo and P. Zhang. *Ambulatory Electrocardiography*. People's Medical Publishing House Co., LTD.

[29] M. Hadjem and F. Naït-Abdesselam. An ECG T-wave Anomalies Detection Using a Lightweight Classification Model for Wireless Body Sensors. In *Workshop on ICT-enabled services and technologies for eHealth and Ambient Assisted Living*, 2015.

[30] B. Hales, M. Terblanche, R. Fowler, and W. Sibbald. Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care*, 20(1):22–30, Feb. 2008.

[31] B. M. Hales and P. J. Pronovost. The checklist - a tool for error management and performance improvement. *Journal of Critical Care*, 21(3):231–235, Sept. 2006.

[32] M. A. Hanson, H. C. Powell, A. T. Barth, K. Ringgenberg, B. H. Calhoun, J. H. Aylor, and J. Lach. Body Area Sensor Networks: Challenges and Opportunities. *Computer*, 42(1):58–65, Jan. 2009.

[33] R. A. Harrigan, T. C. Chan, and W. J. Brady. Electrocardiographic Electrode Misplacement, Misconnection, and Artifact. *The Journal of Emergency Medicine*, 43(6):1038–1044, Dec. 2012.

[34] L. N. Heuvel, D. K. Lorenzo, L. O. Jackson, W. E. Hanson, J. J. Rooney, and

D. A. Walker. *Root cause analysis handbook: A guide to efficient and effective incident investigation*. Rothstein Publishing, 2008.

[35] G. R. J. Hockey and J. Sauer. Cognitive Fatigue and Complex Decision Making Under Prolonged Isolation and Confinement. *Advances in space biology and medicine*, 5:309–330, 1996.

[36] A. R. Houghton and D. Gray. *Making Sense of the ECG, Third Edition*. Hodder Education, Jan. 2012.

[37] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A Practical Guide to Support Vector Classification. Technical report, Department of Computer Science, National Taiwan University, May 2016.

[38] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati. Classification of ECG signals using machine learning techniques: A survey. In *2015 International Conference on Advances in Computer Engineering and Applications*, pages 714–721, Mar. 2015.

[39] E. Keogh, J. Lin, and A. Fu. HOT SAX: efficiently finding the most unusual time series subsequence. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 8 pp.–233, Nov. 2005.

[40] Y. Kishimoto, Y. Kutsuna, and K. Oguri. Detecting Motion Artifact ECG Noise During Sleeping by Means of a Tri-axis Accelerometer. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 2669–2672. IEEE, Aug. 2007.

[41] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190, Nov. 2006.

[42] X. Lai, Q. Liu, X. Wei, W. Wang, G. Zhou, and G. Han. A survey of body sensor networks. *Sensors (Basel, Switzerland)*, 13(5):5406–5447, 2013.

[43] E. A. Lee. Cyber-Physical Systems - Are Computing Foundations Adequate? *Position Paper for NSF Workshop on Cyber-Physical Systems: Research Motivation, Techniques and Roadmap, Austin, Texas*, Oct. 2006.

[44] I. Lee, G. J. Pappas, R. Cleaveland, J. Hatcliff, B. H. Krogh, P. Lee, H. Rubin, and L. Sha. High-confidence medical device software and systems. *Computer*, 39(4):33–38, Apr. 2006.

[45] I. Lee, O. Sokolsky, S. Chen, J. Hatcliff, E. Jee, B. Kim, A. King, M. Mullen-Fortino, S. Park, A. Roederer, and K. K. Venkatasubramanian. Challenges and Research Directions in Medical Cyber Physical Systems. *Proceedings of the IEEE*, 100(1):75–90, Jan. 2012.

[46] W. S. Lee, D. L. Grosh, F. A. Tillman, and C. H. Lie. Fault Tree Analysis, Methods, and Applications : A Review. *IEEE Transactions on Reliability*, R-34(3):194–203, Aug. 1985.

[47] M. Li, W. Lou, and K. Ren. Data security and privacy in wireless body area networks. *IEEE Wireless Communications*, 17(1):51–58, Feb. 2010.

[48] T. H. Lim. *Dependable Network Protocols in Wireless Sensor Networks*. PhD thesis, University of York, 2013.

[49] C.-H. Lin. Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier. *Computers & Mathematics with Applications*, 55(4):680–690, Feb. 2008.

[50] G. Lu, J.-S. Brittain, P. Holland, J. Yianni, A. L. Green, J. F. Stein, T. Z. Aziz, and S. Wang. Removing ECG noise from surface EMG signals using adaptive filtering. *Neuroscience Letters*, 462(1):14–19, Sept. 2009.

[51] X. Lu and Y. Lu. *Xin Dian Tu Ru Men*. Man of Science and Technology, July 2013.

[52] R. Lynch. ECG lead misplacement: A brief review of limb lead misplacement. *African Journal of Emergency Medicine*, 4(3):130–139, Sept. 2014.

[53] S. Z. Mahmoodabadi, A. Ahmadian, M. D. Abolhasani, M. Eslami, and J. H. Bidgoli. ECG Feature Extraction Based on Multiresolution Wavelet Transform. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 3902–3905. IEEE, 2005.

[54] G. Mao, B. Fidan, and B. D. O. Anderson. Wireless sensor network localization techniques. *Computer Networks*, 51(10):2529–2553, July 2007.

[55] D. Meyer, F. Leisch, and K. Hornik. The support vector machine under test. *Neurocomputing*, 55(1-2):169–186, Sept. 2003.

[56] T. Mitchell. *Machine Learning*. McGraw Hill, 1997.

[57] G. B. Moody and R. G. Mark. The impact of the MIT-BIH Arrhythmia Database, May 2001.

[58] G. B. Moody, W. E. Muldrow, and R. G. Mark. A noise stress test for arrhythmia detectors. *Computers in Cardiology*, (11):381–384, 1984.

[59] Y. Mou. *Lin Chuang Xin Dian Tu Jing Jie*. Peking University Medical Press Medical Education, May 2012.

[60] P. L. Myers. Commercial Aircraft Electronic Checklists: Benefits and Challenges (Literature Review). *International Journal of Aviation, Aeronautics, and Aerospace*, 3(1), 2016.

[61] S. Nandi, H. A. Toliyat, and X. Li. Condition Monitoring and Fault Diagnosis of Electrical Motors: A Review. *IEEE Transactions on Energy Conversion*, 20(4):719–729, Dec. 2005.

[62] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt. Cross-Level Sensor Network Simulation with COOJA. In *Proceedings of the 31st IEEE Conference on Local Computer Networks*, pages 641–648. IEEE, Nov. 2006.

[63] J. Pan and W. J. Tompkins. A Real-Time QRS Detection Algorithm. BME-32(3):230–236, Mar. 1985.

[64] K.-J. Park, R. Zheng, and X. Liu. Cyber-physical systems: Milestones and research challenges. *Computer Communications*, 36(1):1–7, Dec. 2012.

[65] G. P. Pizzuti, S. Cifaldi, and G. Nolfe. Digital sampling rate and ECG analysis. *Journal of Biomedical Engineering*, 7(3):247–250, July 1985.

[66] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic. Cyber-physical Systems: The Next Computing Revolution. In *Proceedings of the 47th Design Automation Conference*, DAC '10, pages 731–736, New York, NY, USA, 2010. ACM.

[67] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. 5:1541–1546, July 2005.

[68] Realtime Technologies Ltd. *Shimmer ECG User Guide, Revision 1.11*, 2016.

[69] M. S. Redfern, R. E. Hughes, and D. B. Chaffin. High-pass filtering to remove electrocardiographic interference from torso EMG recordings. *Clinical Biomechanics*, 8(1):44–48, Jan. 1993.

[70] J. J. Rooney and L. N. Heuvel. Root cause analysis for beginners. *Quality progress*, 37(7):45–56, 2004.

[71] A. V. Rosen, S. Koppikar, C. Shaw, and A. Baranchuk. Common ECG Lead Placement Errors. Part I: Limb Lead Reversals. *International Journal of Medical Students*, 2(3):92–98, Oct. 2014.

[72] A. Rudiger, J. P. Hellermann, R. Mukherjee, F. Follath, and J. Turina. Electrocardiographic artifacts due to electrode misplacement and their frequency in different clinical settings. *The American Journal of Emergency Medicine*, 25(2):174–178, Feb. 2007.

[73] A.-B. M. Salem, K. Revett, and E.-S. A. El-Dahshan. Machine learning in electrocardiogram diagnosis. In *2009 International Multiconference on Computer Science and Information Technology*, pages 429–433. IEEE, Oct. 2009.

[74] J. B. Sexton, E. J. Thomas, and R. L. Helmreich. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. *BMJ*, 320(7237):745–749, Mar. 2000.

[75] V. Shnayder, B.-r. Chen, K. Lorincz, T. R. F. Fulford-Jones, and M. Welsh. Sensor Networks for Medical Care. Technical report, Division of Engineering and Applied Sciences, Harvard University, 2005.

[76] S. Q. Simpson, D. A. Peterson, and A. R. O'Brien-Ladner. Development and implementation of an ICU quality improvement checklist. *AACN advanced critical care*, 18(2):183–189, 2007.

[77] The Society For Cardiological Science & Technology. *Clinical Guidelines By Consensus: Recording A Standard 12-lead Electrocardiogram, An Approved Methodology*, Feb. 2013.

[78] K. Soundarapandian and M. Berarducci. Analog Front-End Design for ECG Systems Using Delta-Sigma ADCs. Technical report, Apr. 2010.

[79] J. A. Stankovic, I. Lee, A. Mok, and R. Rajkumar. Opportunities and obligations for physical computing systems. volume 38, pages 23–31, Nov. 2005.

[80] A. Stevenson, editor. *Oxford Dictionary of English (3rd Edition)*. Oxford University Press, 2010.

[81] J. Sun, Y. Fang, and X. Zhu. Privacy and emergency response in e-healthcare leveraging wireless body sensor networks. *IEEE Wireless Communications*, 17(1):66–73, Feb. 2010.

[82] J. Tang. Hardware implementations of multi-layer feedforward neural networks and error backpropagation using 8-bit PIC microcontrollers. In *IEE Colloquium on Neural and Fuzzy Systems: Design, Hardware and Applications*, pages 1–5. IEEE, May 1997.

[83] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman. Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor. In *2007 11th IEEE International Symposium on Wearable Computers,*, pages 1–4, Washington, DC, USA, Oct. 2007. IEEE.

[84] M. Vijayavanan, V. Rathikarani, and P. Dhanalakshmi. Automatic Classification of ECG Signal for Heart Disease Diagnosis using morphological features. *International Journal of Computer Science & Engineering Technology*, 5, Apr. 2014.

[85] Y.-M. Wang, C. Verbowski, J. Dunagan, Y. Chen, H. Wang, C. Yuan, and Z. Zhang. STRIDER: A Black-box, State-based Approach to Change and Configuration Management and Support. Oct. 2003.

[86] A. Whitaker, R. S. Cox, and S. D. Gribble. Configuration Debugging As Search: Finding the Needle in the Haystack. In *Proceedings of the 6th Conference on Symposium on Opearting Systems Design & Implementation - Vol-*

ume 6, OSDI'04, page 6, Berkeley, CA, USA, 2004. USENIX Association.

[87] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.

[88] A. Wood, G. Virone, T. Dona, Q. Cao, L. Selavo, Y. Wu, L. Fang, Z. He, S. Lin, and J. Stankovic. ALARM-NET: Wireless Sensor Networks for Assited-Living Residential Monitoring. Technical report, Department of Computer Science, University of Virginia, 2006.

[89] C.-C. Yang and Y.-L. Hsu. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors*, 10(8):7772–7788, Aug. 2010.

[90] Y. Yao and N. Jiang. Distributed wireless sensor network localization based on weighted search. *Computer Networks*, 86:57–75, July 2015.

[91] C. Yuan, N. Lao, J. R. Wen, J. Li, Z. Zhang, Y. M. Wang, and W. Y. Ma. Automated Known Problem Diagnosis with Event Traces. *SIGOPS Oper. Syst. Rev.*, 40(4):375–388, Apr. 2006.

[92] Q. Zhao and L. Zhang. ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines. In *2005 International Conference on Neural Networks and Brain*, pages 1089–1092. IEEE, 2005.

[93] Q. Zheng, C. Chen, Z. Li, A. Huang, B. Jiao, X. Duan, and L. Xie. A novel multi-resolution SVM (MR-SVM) algorithm to detect ECG signal anomaly in WE-CARE project. In *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, pages 1–6. IEEE, Feb. 2013.

[94] H. Ziade, R. Ayoubi, and R. Velazco. A Survey on Fault Injection Techniques. *The International Arab Journal of Information Technology*, 1(2), July 2004.