VIRGINIA COMMONWEALTH UNIVERSITY

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2011

PROCESSING AND CLASSIFICATION OF PHYSIOLOGICAL SIGNALS USING WAVELET TRANSFORM AND MACHINE LEARNING ALGORITHMS

Abed Al-Raoof Bsoul
Virginia Commonwealth University

Follow this and additional works at: http://scholarscompass.vcu.edu/etd

 Part of the Computer Sciences Commons

© The Author

Downloaded from
http://scholarscompass.vcu.edu/etd/258

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

School of Engineering

Virginia Commonwealth University

This is to certify that the Dissertation prepared by Abed Al Raoof K. Bsoul entitled PROCESSING AND CLASSIFICATION OF PHYSIOLOGICAL SIGNALS USING WAVELET TRANSFORM AND MACHINE LEARNING ALGORITHMS has been approved by his committee as satisfactory completion of the Dissertation requirement for the degree of Doctor of Philosophy

---

Kayvan Najarian, Ph.D., Committee Chair, Department of Computer Science

---

Krzysztof J. Cios, Ph.D., Chair of Computer Science, School of Engineering

---

Vojislav Kecman, Ph.D., Dept. of Computer Science, School of Engineering

---

Rosalyn S. Hobson, Ph.D., School of Engineering

Kevin R. Ward, M.D., School of Medicine

Rosalyn S. Hobson, Associate Dean of Graduate Studies, School of Engineering

---

Russell D. Jamison, Ph.D., Dean, School of Engineering

F. Douglas Boudinot, Ph.D., Dean of the School of Graduate Studies

---

Date

© Abed Al Raoof K. Bsoul, 2011

All Rights Reserved

Dedication

First, I thank my adviser, Dr. Kayvan Najarian, for his guidance and encouragement during my study, which could not have been completed without his support. I am also grateful to all my committee members for their valuable feedback on my work. They are: Prof. Krzysztof Cios, Dr. Rosalyn Hobson and Dr. Vojislav Kecman in the VCU School of Engineering, Dr. Kevin Ward in the MCV Department of Emergency Medicine.

To all my friends and the members of the VCU Biomedical Signal and Image Processing Lab: many thanks for making the route to a Ph.D. a lot more fun.

To my parents, Khaled and Seham Bsoul, who taught me the value of education. I am deeply indebted to them for their continued support and unwavering faith in me. No words can fill these lines to express their unconditional love, support and tranquility throughout my studies.

I would also like to express my gratitude to all my family especially, Rafat, Manal, Samer, Amal, Fidaa and Sanaa for their extended support. With your help in countless ways it was possible for me to complete this research project.

I am also thankful for my son, Khaled for being the best new-born at a critical time.

Most of all, I am eternally grateful for my wife, Esraa for here constant love and strength throughout the years. Without Esraa, and here ability to raise my spirits when I was most discouraged, I could never made it this far. My endless love, you were the wind beneath my wing.

PROCESSING AND CLASSIFICATION OF PHYSIOLOGICAL SIGNALS USING WAVELET TRANSFORM AND MACHINE LEARNING ALGORITHMS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

ABED AL RAOOF K. BSOUL
Master, Yarmouk University (Irbed, Jordan), 2004

Director: KAYVAN NAJARIAN
ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE

Virginia Commonwealth University
Richmond, Virginia
April, 2011

Table of Contents

List of Tables

List of Figures

Abstract

PROCESSING AND CLASSIFICATION OF PHYSIOLOGICAL SIGNALS USING WAVELET TRANSFORM AND MACHINE LEARNING ALGORITHMS

By Abed Al Raoof K. Bsoul, Ph.D.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2011

Major Director: Kayvan Najarian
Associate Professor, Department of Computer Science

Over the last century, physiological signals have been broadly analyzed and processed not only to assess the function of the human physiology, but also to better diagnose illnesses or injuries and provide treatment options for patients. In particular, Electrocardiogram (ECG), blood pressure (BP) and impedance are among the most important biomedical signals processed and analyzed. The majority of studies that utilize these signals attempt to diagnose important irregularities such as arrhythmia or blood loss by processing one of these signals. However, the relationship between them is not yet fully studied using computational methods.

Therefore, a system that extract and combine features from all physiological signals representative of states such as arrhythmia and loss of blood volume to predict the presence and the severity of such complications is of paramount importance for care givers. This will not only enhance diagnostic methods, but also enable physicians to make more accurate decisions; thereby the overall quality of care provided to patients will improve significantly.

In the first part of the dissertation, analysis and processing of ECG signal to detect the most important waves i.e. P, QRS, and T, are described. A wavelet-based method is implemented to facilitate and enhance the detection process. The method not only provides high detection accuracy, but also efficient in regards to memory and execution time. In addition, the method is robust against noise and baseline drift, as supported by the results.

The second part outlines a method that extract features from ECG signal in order to classify and predict the severity of arrhythmia. Arrhythmia can be life-threatening or benign. Several methods exist to detect abnormal heartbeats. However, a clear criterion to identify whether the detected arrhythmia is malignant or benign still an open problem. The method discussed in this dissertation will address a novel solution to this important issue.

In the third part, a classification model that predicts the severity of loss of blood volume by incorporating multiple physiological signals is elaborated. The features are extracted in time and frequency domains after transforming the signals with Wavelet Transformation (WT). The results support the desirable reliability and accuracy of the system.

Novelty and Contributions

As death and complications associated with arrhythmias and severe loss of blood, in particular the ones resulted from traumatic injuries, are increasing, so is the need for computer systems to provide care givers with more detailed and more accurate information about the health of patients. By creating such systems, diagnostic procedures, treatment outcomes and resource management can be improved, and thereby ameliorate the survival rate only by means of quality, but also in a cost effective sense.

Prior work has most often used single physiological signals at a time to detect arrhythmia or predict the severity of loss of blood volume. However, a combination of the signals must also be taken into account. Specific features such as the number of ectopic beats, heart rate and temperature may have an important impact on the outcome and prediction of the severity of arrhythmia, and thereby increase the accuracy of such predictive systems.

This dissertation focuses on the processing of ECG signal to detect the deflection waves to create a model to predict the severity arrhythmia and blood volume loss using wavelet-based methods and machine learning algorithms. The overall study aims to create a system that is capable of combining features - including but not limited to the number of ectopic beats and heart rate - to accurately predict patient's condition from vital physiological signals such as ECG, blood pressure and impedance.

The motivation for focusing on arrhythmia and hemorrhage is the fact that they are broadly prevalent. The majority of the population experience arrhythmias daily whereas severe loss of blood is found only among the spectrum of injuries. The following key components can summarize the systems created in this dissertation:

1) A novel method to detect P, QRS, and T components of electrocardiogram signal, using wavelet-based methods.

Since ECG is one of the best indications on how a human heart functions, analysis and processing of ECG are conducted using many different methods. Generally speaking, since P and T waves are hard to detect, the majority of previous studies concentrate on the detection of only QRS complex. However, the method in this dissertation attempts to detect P and T deflection points along with QRS components. The exact shape of an ECG beat is also important and can cause complications in the detection process; for example, the polarity of the beat can be positive or negative, which greatly affect the detection process. The specified algorithm in this dissertation can be used to extract features from not only the five characteristic waves, but also the overall shape of a beat. This has a major effect on the accuracy of the overall system, in particular on the classification of arrhythmia as well as the prediction of arrhythmia severity procedures.

2) A classification model using Support Vector Machine (SVM) that incorporates wavelet-based features from ECG to classify the arrhythmia associated in the signal and detect severity.

Many algorithms using ECG morphological information are developed to detect abnormal (ectopic) heart beats. However, the extracted set of features, do not include information from P and T waves since they are not easy to detect. Detecting and extracting informative features, can be thought of as the major challenge in creating an arrhythmia detection model. The irregularities caused by arrhythmias affect the shape and timing of P and T waves as well as QRS complex. Therefore, attributes such as the shape of P and T wave, duration of important intervals such as ST-segment and relative amplitude of R wave must be measured

and included in the feature set in order to accurately detect ectopic beats.

The crux of the arrhythmia model implemented in this dissertation is to detect the severity of from ECG signal. The method extracts features from P and T waves, QRS-complex, and interval durations of waves in ECG using a novel detection algorithm. The set of attributes are then fed into the SVM algorithm to train a model, it is then used to classify unknown beats. The output of the model is a vector that contains the labels of each beat. Then, the vector is used as input to a deterministic finite state automaton (DFA) to predict the severity of arrhythmia based on rules provided from expert physicians.

The model is used to extract features, such as the rate of abnormal beats, to predict the severity of blood loss.

3) A prediction method for loss of blood volume severity.

The main goal of this dissertation is to create a computerized model that can predict the severity of blood loss in a scale of three levels; mild, moderate and severe.

Bleeding seems to affect vital biomedical signals such as ECG, blood pressure and impedance. Analysis and processing of such signals are carried out in this dissertation to extract novel features from time domain and wavelet coefficients. In addition, arrhythmia-related features such as the rate of ectopic beats are also extracted and combined to the computed set of features from the aforementioned signals for model building. To evaluate the importance of incorporating multiple physiological signals, several models are created.

The overall system that combines these methods can be used to predict the severity of blood volume loss. The gamut of the systems discussed hereafter provide an insight in creating an

automated decision support system that provides caregivers with accurate information, especially in high-paced environments.

CHAPTER 1 Introduction

1.1 Traumatic Injuries

In the United States, as in much of the world, traumatic injuries are considered the main cause of death and disability in both civilian and combat settings [7, 39, 73, 103, 107]. Almost 36% (41.9 million out of 115.3 million) of all visits made to hospital emergency departments in the United States are injury-related visits [94, 95]. Motor vehicle collision (MVC) is the leading cause of injury death and the fourth leading cause of non-traumatic injuries treated in emergency rooms (ERs) in the United States. More than 39% (343,570 out of 864,736) of deaths from unintentional injuries between 1999 and 2006 are caused by MVCs [38, 124].

One of the most prevalent traumatic injuries is traumatic brain injury (TBI). The Federal Centers for Disease Control and Prevention (CDC) reported in 2004 that 1.4 million TBIs occur in the US annually. Approximately, more than 85% of brain injuries that occur in the US each year are considered mild. Although nonsevere brain injuries do not affect life expectancy, young patients may face several years of neurological disorder (e.g. congenital malformation) leading to substantial inability and unemployment [6, 31, 36, 78]. About 10% to 20% of traumatic injury patients may suffer from post-traumatic stress disorder (PTSD) and 9% to 15% develop major depressive disorder [94, 95]. On the other hand, pelvic injuries caused by high-energy impacts that destroy the integrity of the pelvic ring are associated with a mortality rate of between 5 and 20% [108].

A presumably lethal consequence of traumatic injury, hemorrhagic shock (HS) is defined as the excessive loss of blood due to bleeding [46, 62, 82]. HS may happen internally (inside the body) or

externally, resulting in lack of blood flow throughout the body, hence inadequate tissue perfusion, specifically in major organs such as the heart, lungs, brain, kidneys, etc [4, 46, 62, 82, 106]. Approximately 50% of deaths occurred in battlefields and 80% of civilian trauma mortalities in the United States are as a result of severe loss of blood [85]. Notably, most deaths from pelvic fracture are due to complications other than the fracture itself [34]. As many vital organs are located within the pelvic structure, internal hemorrhage is a particularly high risk complication. In such cases the patient may die from the hemorrhage itself, via exsanguination or shock. Alternatively, death might be due to other conditions caused by hemorrhage such as severe infection [8].

It is reported that a better pre-hospital injury care can decrease both the incidents of trauma life loss during the first few hours after injury and the long-term death and sickness, because such better care may prevent incurable changes that could otherwise lead to death [30, 105]. A major finding in a retrospective study by the Israel Defense Forces is that more than 83% (83 out of 101) of the soldiers were killed during the first hour of injury and hemorrhage accounts for 50% of deaths which makes it the major cause of death in the study group [112]. Another important study finds that over 56% (643 out of 1130) of US military casualties between March and September of 2004 are as a result of battle injuries [126].

An important non-invasive signal that measures the heart activity is the heart rate variability (HRV or RRI). HRV can be obtained from ECG signal when applying a simple QRS detection algorithm. The difference in time between successive R peaks is then calculated to form RRI signal. Recently, analysis of HRV from electrocardiography (ECG) recording has become a popular method for assessing activity of the autonomic nervous system.

Several studies over the past 20 years have discovered significant relationship between HRV and heart diseases [23, 24, 37, 80, 100, 110]. Monitoring irregularity of heart beats observed in

HRV appears to provide insightful information about the heart and neurological diseases as well as physical and mental stress. Specifically, RRI is wildly studied as a measure of cardiovascular function that can be used in risk estimation and diagnosis of cardiac events.

Power spectrum density (PSD) and fractal dimension (FD) are the traditional techniques used for HRV analysis. Previous studies demonstrate that PSD and FD are useful tools to study the heart functionality [24, 28, 29, 37, 80, 100, 110]. A common observation indicates that HRV complexity is reduced in cardiac diseases. However, those traditional methods are limited to the analysis of signals whose statistical characteristics change slowly. Moreover, PSD, naturally, considers that the signal to analyze is at least weakly stationary. However, HRV signal is extremely nonstationary [10, 84, 120]. Analysis and interpretation of nonstationary signals are significantly more complicated than the stationary ones, requiring methodologies such as wavelet transform.

During pre-hospital transportation of traumatic injuries, vital biomedical signals such as ECG and BP are regularly measured to manually assess the patient's condition. Physicians face a challenging task: to integrate the information and make rapid and accurate treatment decisions in a high-pace and stressful environment. Since inaccurate diagnosis severely impacts the welfare of the patient and the cost of treatment, developing a system to automatically assess the severity of injury as early as possible could both improve patient care and reduce costs; thereby enhance the treatment provided by care givers [39]. Currently, however, there is no widely used system that can evaluate the severity of injury by integrating and processing multiple physiological signals. Such a system would need to analyze and process multiple physiological signals and extract informative features to report the latent infirmity of a critical subject.

1.2 Significance of this Study

The clinical significance of the prediction of loss of blood volume system lies in its capability to provide physicians, and care givers with time-sensitive information which allow them to deliver high quality care by improving diagnostic procedures, and making more accurate decisions in high-pace critical environments. In addition, an accurate prediction system may be used in rural and remote areas where trauma experts may not be available.

In the biomedical field, the analysis of ECG signals for the detection of P-QRS-T wave components has received considerable attention. Wavelet transformation has previously been used to detect all deflection points of ECG signal. In this dissertation, the detection system incorporates a computationally efficient method compared to discrete wavelet transform which is called dual-tree complex wavelet transform (DT-CWT). The algorithm is not limited to be used in prediction of bleeding severity or arrhythmia detection systems. Any study that manipulates the ECG signal to discover similar diseases and complications may benefit from the capability of this system.

1.3 Aims of this Study

The ultimate aim of this study is to construct a lucid system that can discover the latent risk associated with excess loss of blood volume by integrating relevant knowledge from biomedical signals, to support health care providers in diagnosis and treatment of injuries. The specific aims for this study are:

1) Analyze and process electrocardiogram signals using dual-tree complex wavelet transform (DT-CWT) to delineate the main deflection points (i.e. P-QRS-T).

2) Create a classification model using support vector machine learning algorithm to discriminate

between three functional types of heartbeats; normal (N), premature atrial contraction (PAC) and ventricular ectopic beat (VEB).

3) Amalgamate features from vital physiological signals (i.e. ECG, BP and impedance) and arrhythmias (e.g. number of abnormal beats) to predict the severity of loss of blood volume in patients with injury.

The remaining chapters of this dissertation are organized as follow: Chapter 2 describes the related work as well as the advantages and limitations of the study. Chapter 3 presents the detection algorithm of ECG deflection points. Chapter 4 explains the methodology for arrhythmia classification and severity detection. Chapters 5 describes the system for the severity of blood loss prediction. Chapter 6 contains an evaluation of the systems implemented. Chapter 7 describes the time complexity of the algorithms. Chapter 8 elaborates the conclusions of the study as well as the future work.

CHAPTER 2 Background and Related Work

This chapter elaborates previous work related to methods addressed in this study. First, a discussion of previous studies about the detection of ECG characteristic points is presented. Algorithms specific to arrhythmia detection from ECG signals are then reviewed. Finally, methods specific to detect and predict loss of blood volume severity are surveyed.

2.1 ECG Detection Systems

During the last decade, several ECG detection methods have been proposed. The majority of these systems focus only on the detection of QRS-complex, since the relative magnitude of R wave is much higher than the other waves (P and T). As such, studies that detect P and T components along with QRS-complex are sparse. Meanwhile, different methods for QRS detection have been used in several applications; for instance, approaches based on signal derivative [15, 17, 48, 63, 99], mathematical morphology [117, 121, 133, 134], Hidden Markov Models (HMM) [20, 21, 44], Hilbert Transform (HT) [9, 11, 42], Artificial Neural Network (ANN) [97, 119, 131], and Wavelet Transform (WT) [75, 79, 83, 109] are among the most popular studies.

Derivative methods are based on calculating the first or second derivative of the signal to enhance the slope of the R wave. Although noise is the main challenge for such methods. It can be reduced by applying a bandpass filter as a preprocessing step. Derivative-based algorithms are computationally simple and are understandable.

Mathematical morphology approaches are based on the terms erosion ($\ominus$) and dilation ($\oplus$) that are drawn from image processing field, and applied for ECG signal-to-nose enhancement.

Derived operators are opening ($\circ$) and closing ($\bullet$), that are designed as signal morphologic filters. The detection of QRS is done by applying a sequence of mathematical morphology operations. In addition noise is reduced and baseline contaminations are removed by successive application of the morphological operators. The results of mathematical morphology are better than derivative methods. However, the use of threshold values are selected empirically which makes it hard to create a general method for all types of datasets.

HMM methods use a probability function that changes according to the underlying Markov chains to model the observed data sequence. The crux of these algorithms is to extract the underlying state from any given portion of a signal; for example in ECG signals, possible states are P wave, QRS, and T wave. The main drawback for such methods are the need to manually segment each ECG cycle before training.

Hilbert transform approaches utilize the fact that Hilbert transform is an odd function. This means that Hilbert transform of a signal will result in another signal that will cross zero on the time axis (x-axis) whenever there is a deflection in the original signal. Though, noise and baseline drifts will influence the transformation. However, it is not a disadvantage since the effect can be reduced by applying a filter to ameliorate signal-to-noise ratio. Although results of Hilbert transform are easy to understand and interpret, calculating Hilbert transform of long signals is very slow even when Fast Fourier Transform (FFT) is used.

ANNs have been broadly used in nonlinear signal processing, classification, and optimization. In ECG signal processing, multilayer perceptron (MLP), radial basis function (RBF) networks, and learning vector quantization (LVQ) networks have been employed. While efficiency of ANNs is better than classical linear methods for several applications. Optimization of the network parameters (i.e. number of neurons, the coefficients, center vectors and the standard deviations of

the bases function) requires a great deal of computational effort.

Of particular relevance to the detection method adapted in this research are studies that apply wavelet transformation. The transform of a signal yields in a time-scale representation similar to the representation of time-frequency of the short-time Fourier transform (STFT). Unlike STFT, WT uses a set of bases functions that allow flexible time and frequency resolution for different frequency bands. The analysis function is called the mother wavelet, which is often a short oscillation, with finite time, and zero mean. Several mother wavelets exist. Unfortunately, there is no universal rule to choose among them. However, the mother wavelet to be used is preferred to captures the shape of the dominant waveform in the processed signal [92]. For digital signals, the discrete wavelet transform (DWT) is applied. DWT can be easily implemented on digital signal processors which makes it a valuable tool in the biomedical field. Moreover, the computational complexity of DWT is of order $O(n)$, which means it is very fast and simple.

The detection method in this study incorporates dual-tree complex wavelet transform (DT-CWT), which is a recent enhancement to DWT. The importance of using this mother wavelet are two fold; it is shift invariant, and is based on computationally efficient filter back [113].

2.2 Arrhythmia Detection Systems

Since it is a potentially dangerous and life threatening complication, considerable attention has been devoted to the field of arrhythmia detection. Arrhythmia is any variation in the pacemaker sites or any delay in the conduction network of the heart [127]. Because ECG reflects the functionality of the heart, many algorithms have been proposed for ECG beat classification to detect arrhythmias. Almost all methods that process ECG to detect arrhythmia have two stages: feature extraction and beat classification.

In arrhythmia classification from ECG, the types of features are of paramount importance. Attributes from ECG timing [28, 87, 98], higher order cumulant [35, 97], multiscale morphological derivative (MMD) transform [118], wavelet transform (WT) or Fourier transform [2, 45, 53, 116, 123] are among the different types considered.

Approaches that compute features from ECG morphology (e.g. QRS width) are easy to understand and do not need extensive calculations. Although this type of feature is sensitive to variations and noise. A suitable filter could be researched to improve signal-to-noise ratio. Higher order cumulant features are less susceptible to the aforementioned changes. Nevertheless, they are computationally expensive. The inherent noise in a single is reduced when features from MMD transform are extracted. However, the transform must be defined at different scales to determine the best representative scale of the signal. Frequency domain features by signal transformation such as Fourier and WT are robust to variations in the signal. Indeed, WT is better than Fourier transform for ECG analysis, since it is non-stationary [92].

For classification, the methods used include linear discriminants [28], artificial neural networks (ANNs) [35, 45, 97, 98], autoregressive modeling (AR) [41], Hidden Markov models (HMM) [2] and support vector machines (SVM) [53, 87, 116, 123].

Good results can be achieved when linear discriminant methods are considered. However, much of the work is spent to find the number of discriminant variables that elicit the best separation between the classes. Moreover, they are time consuming, hence to implement them in real-time applications is neither simple nor easy. Although, AR models are clear and easy to understand, they assume linear relationship between the samples of the signal. However, ECG beats are non-linear by nature [92]. The use of ANNs for beat classification is capable of providing more accurate results than the aforementioned models. However, the cost function of neural networks is

to minimize the sum of square between the decision function and the training data points, though the accuracy is very sensitive to the overlapping between the data points. In addition, ANN methods can trap into local and global minima. HMMs algorithms provide comparable results to ANN algorithms, but they exploit large assumptions about the data and the number of parameters that will be used.

Arrhythmia classification studies which incorporate support vector machines (SVM) are relevant to the work in this dissertation. The key features of SVMs are two folds; the maximization of the margin between the closest examples from each class and a few samples are involved in the determination of the classification (sparse solution). This feature gives the superiority of SVM over ANN. The main challenge when using SVM is the choice of the kernel.

2.3 Existing Models for Loss of Blood Volume

Most deaths as a result of traumatic injury are due to hemorrhage [30, 107]. When loss of blood volume is detected and assessed as early as possible, incurable consequences may be prevented or avoided. Therefore, injury-related death may be reduced. However, studies on the assessment of the severity of blood loss are scarce.

Lee et. al. (2010) [74], creates a survival prediction model for rats, to achieve early diagnosis of hemorrhagic shock. The method incorporates heart rate, mean arterial pressure (MAP), respiration rate and temperature. An artificial neural network model is then created to predict the survival rate. The sensitivity and specificity of survival prediction were 98.4 and 96.6%, respectively. The limitation of the study is that the data samples are from rats. Therefore, results are not directly applicable to human physiology.

Hakimzadeh et. al. (2009) [47], study the prediction capability of a non-invasive biomedical

signal called Transcranial Doppler (TCD). The implementation applies advanced signal processing and machine learning methods. The research has two parts; the first one is to categorize the severity of blood volume loss into three functional classes; mild, moderate and severe. The second part considers two classes only; severe and non-severe. The study allows assessment of volume loss and prediction of hemorrhagic shock (HS), particularly in cases of traumatic injury. This method can be used in real-time monitoring of internal bleeding during surgical procedures. Results of the first part shows that the method has 70% accuracy and for the second part, the accuracy is more than 84%.

Ji et. al. (2009) [60], study the capability of low level physiological signals using discrete wavelet transformation (DWT) and machine learning methods to predict the severity of hemorrhage encapsulated in the patterns of these signals. Physiological signals such as ECG, BP and thoracic impedance (IZT and DZT) are used for feature extraction. Machine learning algorithms are used to create a model that can distinguish between three classes as in the study mentioned before; mild, moderate and severe. The dataset is obtained from lower body negative pressure (LBNP) procedure. The research shows that the highest accuracy obtained is 82% when using support vector machine technique.

Porter et al. (2009) [100] study the change of RRI during different stages of hemorrhage in rats. Frequency-based features from power spectrum are calculated. Three types of hemorrhage are recorded (low rate, intermediate rate, and fast rate). The findings of this study show that the heart rate increase while atrial pressure (AP) did not change significantly during 15% of total blood volume loss. Moreover, the study highlights the importance of the rate of blood loss and its effect on the compensation process. Since the study is done on rats, its results are not directly applicable to human physiology.

Chen et al. (2008) [13], formulates a tool that incorporates vital signals to identify patients who require blood transfusion. The study uses two classes; hemorrhagic, patients who received blood in emergency room, and control, patients who did not. The method extract features from ECG, photoplethysmogram, respiratory and patient attribute data. For classification tasks [13] uses a linear classifier that is trained using least-square method. Expressed as the area under a receiver operating characteristic curve (ROC), the accuracy of the classifier is 75%.

Batchinsky et al. (2007) [5] analyze hemorrhage shock in group of sheep. As other studies, HRV is monitored and transformed by Fourier transformation to compute the power spectrum of RRI. In addition to HRV, atrial blood pressure (ABP) is also monitored. The main observations of this study is that the heart rate and systolic arterial pressure (SAP) start to change after 20 minutes of hemorrhage, but the diastolic arterial pressure (DAP) decreased continuously. Again, the study is conducted on animals only.

Cooke et al. (2006) [24], presents an approach that analyzes data from pre-hospital trauma patients to test if the survival rate is correlated to higher parasympathetic and lower sympathetic autonomic activity. In addition to Galsgow coma scale score (GCS) and patients' demographic information, signals such as ECG and atrial blood pressure (ABP) are used. Heart rate variability (HRV) is also calculated by ECG filtering and R peak detection. Ectopic beats are excluded from the study and features from time-domain and frequency-domain are extracted. Specifically, features from power spectrum density (PSD) as computed using Fourier transform. Statistical analysis is carried out to discriminate between surviving or died patients.The assumption about ECG to be ectopic-free and the use of power spectral which assumes the signal is stationary, limit use of results in real application.

In conclusion, traditional physiological signs such as heart rate variability have a limit in de-

tecting and predicting the severity of hemorrhage. Research conducted on animals are not directly applicable for humans. Basic, vital and non-invasive biomedical signals have proved the prediction of hemorrhage, but still real-time applications are very hard to implement using these methods. Since pre-hemorrhage status, the beginning of hemorrhage, and rate of blood loss are generally unknown. Incorporating multiple physiological signals such as ECG, ABP and impedance, in addition to the severity of arrhythmia latent in ECG signal, may have a better chance of creating a system for detection and predicting the severity of hemorrhage.

2.4 Approach of this Study

The novelty of the system described in this dissertation is, its ability to combine features extracted from physiological signals that can be easily acquired to detect and predict the severity of hemorrhage. Unlike other hemorrhage models, the implemented method performs successfully in the presence of noise, baseline drifts and abnormal beats. Moreover, the datasets considered are by applying the lower body negative pressure (LBNP) protocol on humans. It is proved in [22, 23] that LBNP model induces similar physiological responses that simulate acute hemorrhage in humans.

The approach of the study is hierarchical; he method first detects the characteristic points in ECG signal, and uses the information obtained to identify the type of arrhythmia beat-by-beat. Arrhythmia classification model is then applied classify the test dataset in order to predict the severity of arrhythmia using a deterministic finite automaton (DFA). In addition to the information obtained from arrhythmia, raw signals of ABP and impedance are also used to train a model by SVM machine learning method to output the severity of the hemorrhage. More informed decision can be taken by care givers after the severity is estimated and acquired by the system.

The overall process is outlined in Figure 2.1. More detailed schematic diagrams for each

process are provided in relevant chapters.

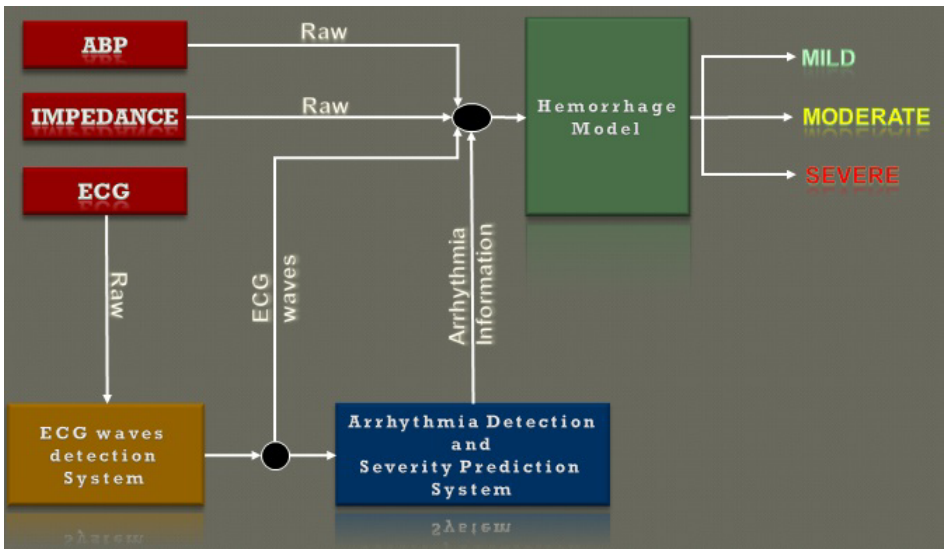


Figure 2.1: Schematic diagram for the systems in this dissertation

CHAPTER 3 Detection of ECG Characteristic Points

This chapter introduces the method for detection of ECG waves by combining the capabilities of dual-tree complex wavelet transform (DT-CWT) to capture the morphological characteristics of ECG signal. The algorithm can be used for ECG processing applications, such as calculation of the HRV signal. The results prove that the method is robust against noise and typical artifacts in ECG signal, such as baseline drift.

The rest of this chapter is organized as follows. Section 3.1 introduces the physiology behind ECG. A detailed description of the algorithm is presented in Section 3.2. Results and their comparison to those of other algorithms are given in Sections 3.3 and 3.4. Section 3.5 Summarize the chapter.

3.1 Introduction

Several clinical details are encapsulated as intervals and amplitudes in ECG signal (see Figure 3.1). Many algorithms have been developed in order to extract such information. The most important clinical details in ECG signal are P, Q, R, S, and T waves. Sometimes, a sixth wave (U) may follow T. The waves Q, R, and S are grouped together to form QRS-complex, which in turn can be used to detect P, and T

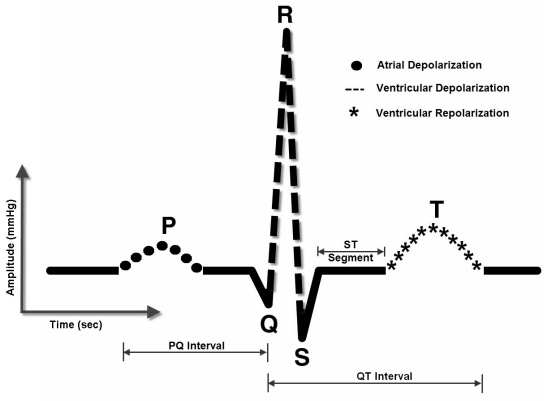


Figure 3.1: One cycle of ECG based upon cardiac physiology. Atria depolarization, ventricular depolarization, and ventricular repolarization are represented as in a normal beat

waves not only because of its central location relative to these waves, but also its relatively high amplitude which makes this complex easier to detect.

When the sinoatrial (SA) node generates an impulse, a depolarization wave moves from the right atrium to the left atrium. If the wave is measured by electrodes, the P wave in ECG is shaped and lasts for about 0.08 seconds. The atria contract about 0.1 seconds after P wave starts. The electrical activity travels from the SA node towards the atrioventricular (AV) node, which will result in ventricular depolarization.

Ventricular contraction is accompanied by depolarization. As a result of ventricular depolarization, the QRS is formed in ECG. The duration of QRS is between 0.08 to 0.12 seconds. Since the mass of the ventricles is greater than the atria, the magnitude of QRS is higher than other waves, such as P and T.

After the ventricles contract, they are repolarized and T wave is generated and lasts about 0.16 seconds. T wave is more prolongated than the other waves such as QRS and P waves because the repolarization is slower than depolarization. While the atria repolarize after they depolarize and since the QRS complex is recorded at the same time as the repolarization of the atria, the resulting wave from the repolarization is obscured.

Literature shows numerous methods for R and QRS detection, however, P and T wave detection methods are rare.

In this dissertation, an algorithm based on DT-CWT for the detection of QRS-complex, and wavelet transformation (WT) using Daubachi 2 (db2) as the mother wavelet at level 2 for the detection of P, and T waves is presented. The method first detects QRS complex, then P wave, followed by T wave. The implemented detection system is used to extract informative features from not only QRS-complex, but also from P and T waves. The extracted features will be used to

classify the severity of blood loss. Figure 3.2 shows the overall process for the method. The next section elaborates the details of the system.

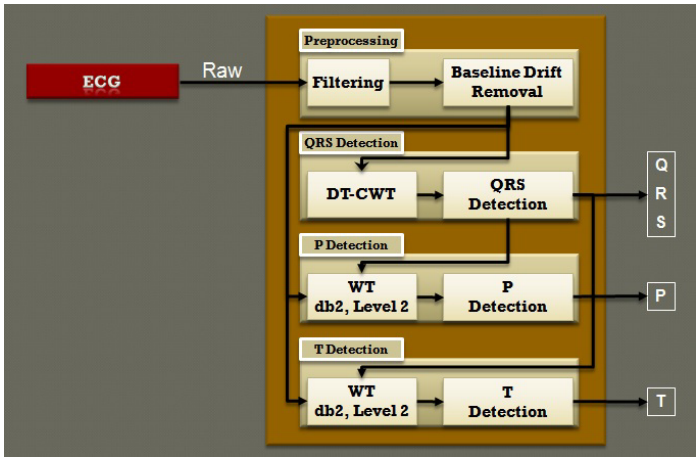


Figure 3.2: Schematic diagram for ECG waves detection system

3.2 Methodology

3.2.1 Data Specification

To ensure the accuracy and reliability of the results, two different dataset are used for testing and validation. The details for each of them is given herein.

3.2.1.1 USAISR LBNP Dataset

In this dataset, 93 subjects from Lower Body Negative Pressure (LBNP) [23] are considered. LBNP is a protocol where the lower half of the body is placed in a depressurized chamber. One purpose of (LBNP) chambers is to simulate the transition from earth gravity to micro-gravity. Physiological tests are conducted to assess stresses upon the cardiovascular system during these simulations. In general, the internal negative air pressure of LBNP chambers is controlled with a proportional control system using only air-pressure as input. In each test, the subject experienced multi-stages air pressure where in each stage, the level of negative air pressure is increased by

a step of $-15mmHg$ for 5 minutes and after 20 minutes, the pressure is decrease by $-10mmHg$, while ECG together with other physiological signals such as ABP and thoracic impedance are sampled at $500Hz$. Table 3.1 shows the LBNP level for each stage.

Table 3.1: Different pressure levels during LBNP procedure

LBNP protocol	Stage	Time
0 mmHg	Baseline	5 Min
-15 mmHg	Stage 1	5 Min
-30 mmHg	Stage 2	5 Min
-45 mmHg	Stage 3	5 Min
-60 mmHg	Stage 4	5 Min
-70 mmHg	Stage 5	5 Min
-80 mmHg	Stage 6	5 Min
-90 mmHg	Stage 7	5 Min
-100 mmHg	Recovery	5 Min

The dataset is provided by the U.S. Army Institute of Surgical Research (USAISR) under a protocol approved by the Institutional Review Boards of both the USAISR and Virginia Commonwealth University (VCU).

3.2.1.2 MIT/BIH Dataset

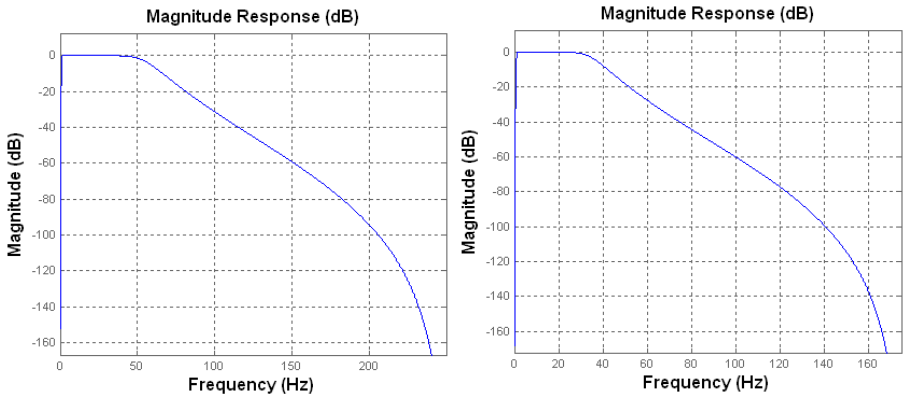
The MIT/BIH arrhythmia database [81] contains 45 half-hour recordings, each containing two ECG lead signals that are band-pass filtered at $0.1 - 100Hz$ and sampled at $360Hz$. There are more than $100,000$ QRS complexes in this database. Some records have a clean R-peaks with few artifacts, while in some other subjects it is difficult to detect the P-QRS-T components due to contamination such as abnormal morphology, noise, and other artifacts in the signal.

3.2.2 Preprocessing of ECG signal

As Figure 3.2 shows, the first step in the detection procedure is ECG preprocessing. This step is intended to remove noise and baseline drifts caused by subject movements or respiration. The output of this step is important for further analysis since it will enhance the signal quality. The step is divided into two main sub-processes, namely, ECG filtering and ECG baseline drift removal, where the output of the first sub-process is used as input to the other.

3.2.2.1 Filtering

Since noise causes different distortion to a signal, a suitable filter must be chosen to remove it [86]. For each dataset in this study, a different bandpass filter is designed. A Butterworth bandpass filter of order 4 and cutoff between 1 Hz and 55 Hz is designed for the LBNP dataset, whereas for the MIT-BIH dataset the cutoff is between 1 Hz and 35 Hz with the same order. After a comprehensive comparative study using Fast Fourier Transform (FFT), these cutoffs proved to reduce the roll-off of the filter. Figure 3.3 shows the overall frequency response of the bandpass filters.

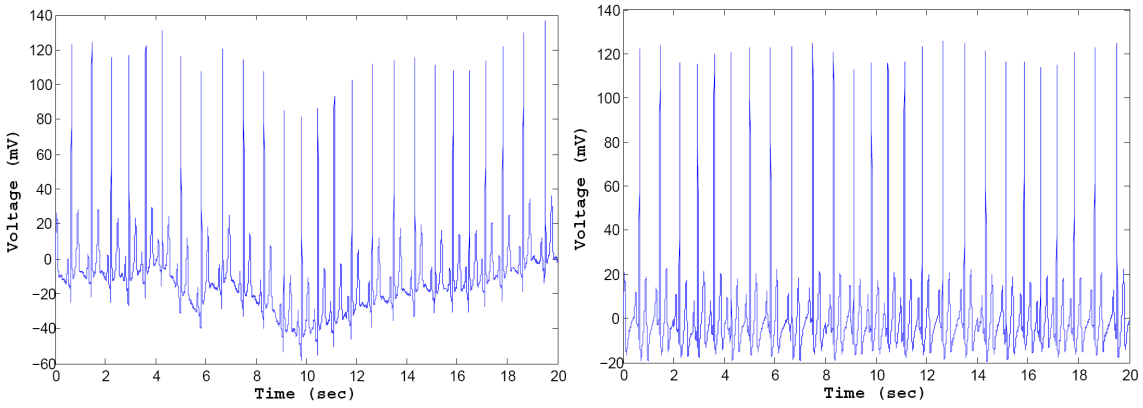


(a) The cutoffs for LBNP are 1 – 55Hz (b) The cutoffs for MIT-BIH are 1 – 35Hz

Figure 3.3: The amplitude response of the digital bandpass (3 dB) used for each dataset

3.2.2.2 Baseline Drift Removal

Any change in the AC current while measuring the ECG causes a drift in the baseline of the signal. Respiration, muscle contraction, and electrode impedance changes due to breathing or movement of the body are the most significant sources of baseline drift in most types of ECG recordings [40]. The presence of baseline drift in ECG signals influences the visual interpretation of the ECG as well as the results obtained from computer-based off-line ECG analysis [125]. For example, the amplitude of a signal at a specific time is harder to obtain when the signal contains baseline drift (see Figure 3.4 (a)).



(a) ECG signal before baseline drift removal

(b) ECG signal after baseline drift removal

Figure 3.4: The effect of baseline drift removal on ECG signals (20 seconds of ECG sample, sampling rate is 500Hz)

In this method, the ECG baseline drift removal is conducted by subtracting the regression line that best fits the samples within a window of size equal to the sampling rate; for example the window size for the LBNP data is 500, whereas the window size for MIH-BIH data is 350. Intuitively, the drift for each second is being removed.

The best fit line associated with n points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ has the form:

$$y = mx + b \tag{3.1}$$

Where y is a point on the line, m is the slop of the line and b is the intercept.

The slope, m, and the intercept of the line, b, in equation 3.1 are calculated as follows

$$m = \frac{n(\sum_1^n xy) - (\sum_1^n x)(\sum_1^n y)}{n(\sum_1^n x^2) - (\sum_1^n x)^2}$$

$$b = \frac{\sum_1^n y - m(\sum_1^n x)}{n}$$

After subtracting the best fit line of the samples in the window, the signal is drift free clean (see Figure 3.4 (b)) and is ready for the next step which is dual-tree complex wavelet transform (DT-CWT).

3.2.3 Wavelet Transformation

The WT of a function $f(t)$ is an integral transform defined by,

$$Wf(a, b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}^*(t)dt. \tag{3.2}$$

where $\psi^*(t)$ denotes the complex conjugate of the wavelet function $\psi(t)_{a,b}$. WT uses a set of basis functions that allows variable time and frequency resolution for different frequency bands. The set of basis functions, the wavelet family $\psi_{a,b}$, is deduced from a mother wavelet $\psi(t)$ as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{2}}.\psi\left(\frac{t-b}{a}\right) \tag{3.3}$$

where a and b are the dilation (scale) and translation parameter, respectively. The mother wavelet is a short oscillation with zero mean.

Discretizing the scale and translation (shift) parameters in (3.3) results in a discrete wavelet transform (DWT). The discretization should obey some defined rules. The usual choice is to follow a dyadic grid on the time-scale plane with: $a = 2^k$ and $b = 2^{kl}\ k \in Z$ [92]. The mother wavelet $\psi(t)$ is then modified to form the basis functions as:

$$\psi_{2^k,b}(t) = \frac{1}{2^{k/2}}.\psi\left(\frac{t-b}{2^k}\right) \tag{3.4}$$

The dyadic DWT (DyWT) integral transform can be obtained by substituting $\psi_{2^k,b}(t)$ from Equation (3.4) in Equation (3.2). Although defined as an integral transform, the DyWT is usually implemented using a dyadic filter bank where the filter coefficients are directly derived from the wavelet function used in the analysis [92]. The decomposition is carried out by cascading two types of filters, Low Pass Filter (LPF) which outputs the approximation coefficients and is followed by down sampling, and High Pass Filter (HPF) which outputs the detailed coefficients and is followed by down sampling.

Figure 3.5 shows level 4 decomposition of a signal. DWT can be applied to ECG signals using this filter bank.

There are many popular mother wavelets $\psi(t)$ including Daubechies wavelets, Mexican Hat wavelets and Morlet wavelets. The Daubechies wavelet family contains the Haar wavelet, i.e. db1, which is the simplest one.

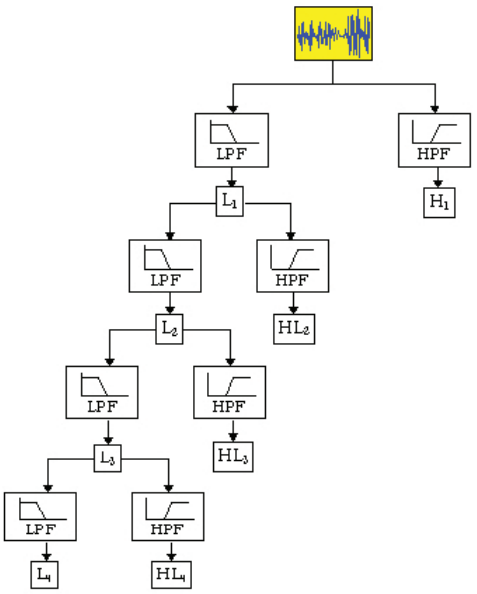


Figure 3.5: Decomposition of a signal with filter bank by cascading LPFs and HPFs

Unfortunately, there is no universal rule to fol-

low when choosing the best mother wavelet for decomposition. However, typically, the mother wavelet should resemble the dominant wave in the signal [92].

When wavelet coefficients are used for real-world signals, they enable near-optimal signal processing based on simple thresholding, denoising, approximation, and deterministic and statistical signal and image algorithms [115].

However, apart from being computationally efficient, the wavelet transform have four fundamental, intertwined drawbacks as follows:

1. OSCILLATIONS

   As mentioned before wavelets are bandpass functions. Therefore, the generated coefficients tend to have large positive and negative oscillations around singularities. In addition, wavelet function crosses through zero several times. As a result, wavelet coefficients could be very small or near zero. This complication makes singularity extraction extremely hard [16].

2. SHIFT VARIANCE

   In addition to positive and negative oscillations around singularities, wavelet coefficients are disturbed by small shifts oscillate around the singularities of the signal. This is another complication of wavelet transform, since the coefficients must be competent to the wide spectrum of patterns exist in the signal [33, 102].

3. ALIASING

   Wavelet coefficients are computed via iterated discrete-time down-sampling operations of

low and high-pass filters that are, of course, not ideal. As a result, substantial aliasing exists in the transformation. This application has this drawback only when the signal needs to be reconstructed [113].

4. LACK OF DIRECTIONALITY

   The product construction of wavelets results in a pattern that is oriented along several directions. Lack of directional influence the processing of multi-dimensional analysis of geometric image, such as edge detection [113].

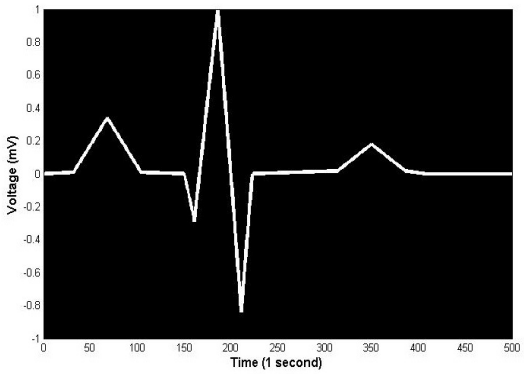
The method implemented in this dissertation will be subjected to the first two complication if wavelet transform is considered for QRS detection. Therefore, DT-CWT at level 4 is used for this purpose.

Figure 3.6 shows the effect of the transformation of a normal ECG beat, when (3.6b) DWT db4 at level 4 is used and (3.6c) DTCWT at level 4 is applied.

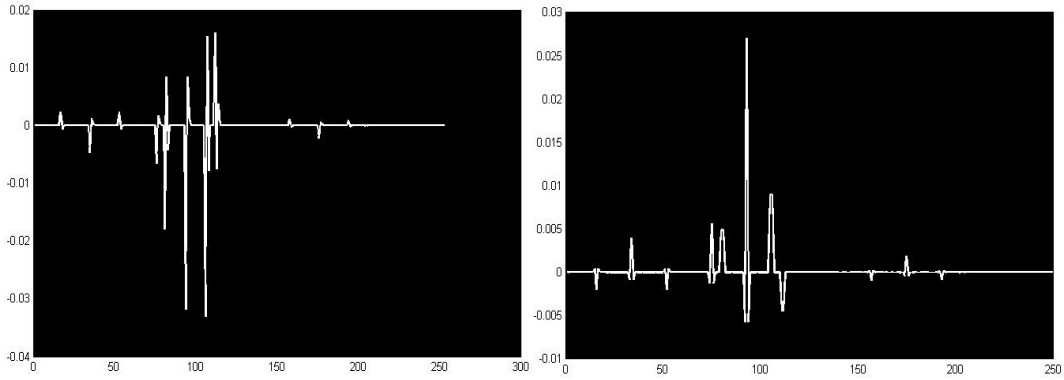
However, since the aforementioned drawbacks have very small effects over the patterns of those two waves, wavelet transform is used in this study to detect P and T waves. Empirically, Daubechie 2 (db2) at level 2 is incorporated for the detection of P and T waves. Figure 3.7 presents the mother wavelet of the decomposition.

3.2.4 QRS Detection

The QRS complex is the most important characteristic waveform of the ECG signal. Its high amplitude makes QRS detection easier than the other waves. Thus, QRS is generally used as a reference within the cardiac cycle [83]. In this study, duration of a cardiac cycle, C_i, as the time interval between $(R_i - R_{i-1})/2$ and $(R_{i+1} - R_i)/2$ where R_i is the time index of the R wave of cycle i is considered.



(a) Normal ECG signal



(b) DWT of db4 at level 4 detail coefficients

(c) DTCWT level 4 detail coefficients

Figure 3.6: Illustration of the transformation of a normal ECG using DWT with db4 and DTCWT at level 4

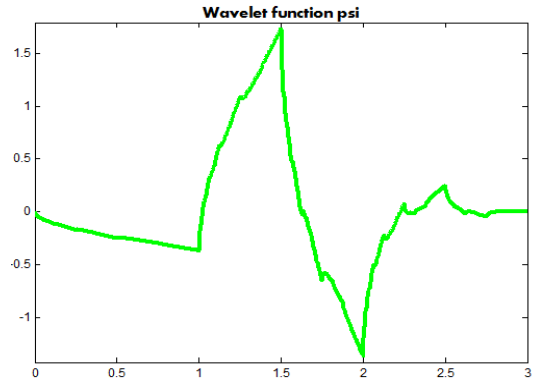


Figure 3.7: The mother wavelets used in the analysis

In order to detect QRS complex, the DT-CWT of ECG at level 4 is calculated. The resulted detail coefficients are then squared to detect the peaks (R waves) in the signal. Figure 3.8 shows the resulted detail coefficients of the transformation when DT-CWT at level 4 is incorporated on 8 seconds ECG signal.

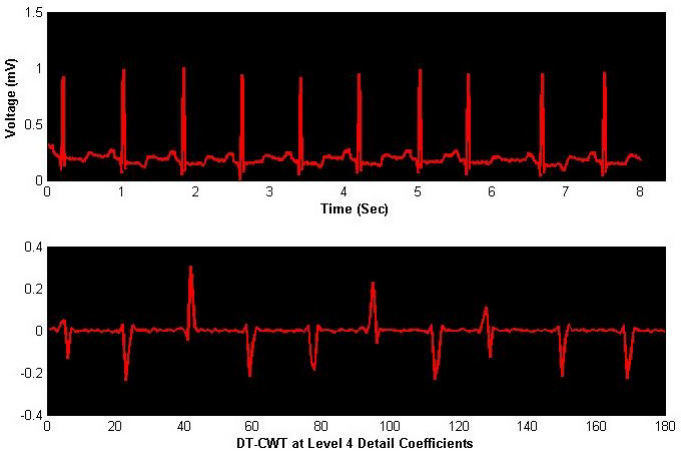


Figure 3.8: ECG signal after DT-CWT is applied on 8 seconds (Sampling rate is 360 Hz). The decomposition is at level 4

A threshold α is applied to the squared detailed coefficient. After examining the data for all subjects in the data set, α is set to $\alpha = 2 \cdot \sigma$ according to variations across these coefficients, where σ is the standard deviation of detailed coefficient. The threshold α gives a measure of the minimum error of detection. Then R_i in each cardiac cycle C_i is detected for $i = 1, 2, 3, ..., n$ where n is the number of cardiac cycles in ECG signal. Figure 3.9 shows the steps of squaring, and thresholding applied to the detailed coefficient.

R_i is obtained from the detailed coefficient. To find the location $L(R_i)$ of cycle C_i. The location obtained from the detailed coefficient (R_i) multiplied by 4. This is because the detailed coefficients are obtained at level 4, which means we downsampled the signal by 4; for example, if the detected R_i from the detailed coefficients is at location k, then $L(R_i)$ is at location $4 \cdot k$ in the original signal.

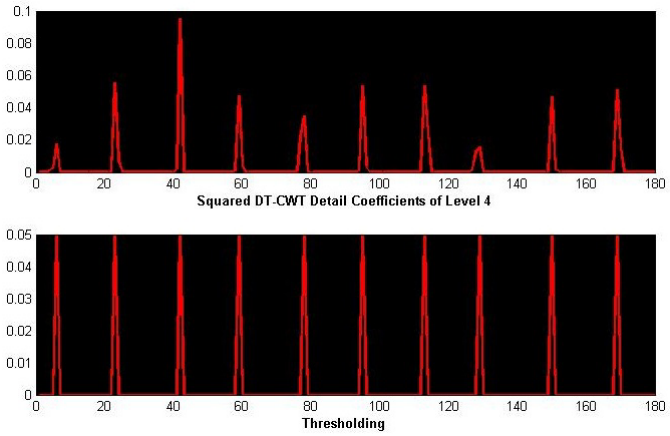


Figure 3.9: The detailed coefficients squared and thresholded by 2 standard deviation

Since the signal may have noise after filtering, a window of size $0.1 \cdot f_s$ centered at $L(R_i)$ is defined, where f_s is the sampling frequency. This window is defined to detect the exact location of R peak in the original signal. In other words, this window is defined to reduce the effect of noise that may still exist in the signal after filtering.

This window size is obtained from the physiological meaning of QRS signal. As mentioned before, the QRS lasts between 80 to 120 milliseconds, and so the expected duration of QRS is 100 ms which is equal to $0.1 \cdot f_s$. The range of the defined window is $[(L(R_i) - 0.1 \cdot f_s), (L(R_i) + 0.1 \cdot f_s)]$

Figure 3.10 shows that QRS complex can have a positive or negative polarity. To distinguish if QRS is either positive or negative, the absolute maximum $|M|$ and absolute minimum $|m|$ amplitudes of the signal in the previously defined window are calculated and compared.

Assume that the maximum amplitude at x_i

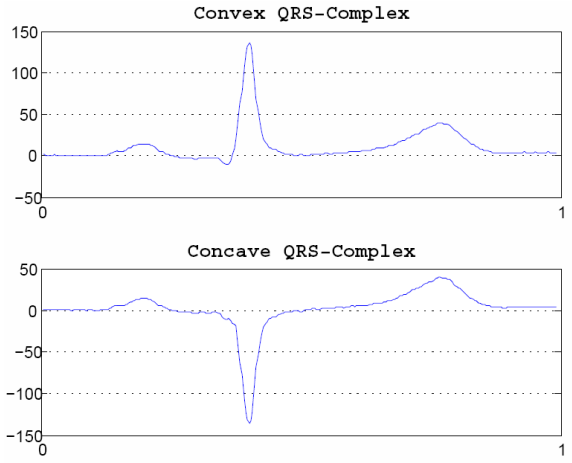


Figure 3.10: Two possible shapes of QRS-complex

is M and the minimum amplitude at x_j is m,

where $x_i \neq x_j$. The decision about the QRS complex polarity is determined by the following:

if $|M| > |m|$ **then**

$L(R_i) \leftarrow i$

else

$L(R_i) \leftarrow j$

end if

The pseudocode means that, if R is positive (convex), then the sample which has the the maximum amplitude in the window is used as the correct location of R wave in the original ECG signal, and if its negative (concave), the position of the minimum is used.

Once R is detected, Q is searched for within a window of size $(0.1 \cdot f_s)/2$ to the left of the detected R. If R is convex, then the location of the minimum magnitude in this window is found and detected as Q, otherwise the location of the maximum amplitude is used.

For S wave detection, the same steps for Q wave are followed with one exception; the window is now to the right of the detected R wave.

3.2.5 P and T Detection

For P wave detection, a window from the beginning of the cycle to the detected Q wave is defined. The same steps for QRS complex detection are taken, but instead of using DT-CWT at level 4, db2 at level 2 is used as the mother wavelet, and the standard deviation ($\alpha = 1$) of the squared detailed coefficient is used as the threshold.

For T wave detection, the same steps for P wave detection are followed, but within a window from the detected S wave to the end of the cycle.

The beginning of the next cycle is obtained by adding $(0.16 \cdot f_s)/2$ to the location of the detected T wave of the previous ECG cycle. This is because the T wave lasts about $0.16 seconds$; the division by 2 is due to the fact that T is approximately located in the center of the wave.

3.3 Results

In this section, the performance of the algorithm over the data sets used is presented. The criterion for reporting results is as follows: if the difference between the detected wave by the algorithm and the actual location is less than $0.02 seconds$, the detection is considered as valid, otherwise it is invalid.

As suggested by literature, the results are presented by the percentage of error, Sensitivity (Se), and positive predictability P^+ according to the following equations:

$$\%Error = \frac{(FP + FN)}{\#Cycles} \tag{3.5}$$

$$Se = \frac{TP}{(TP + FN)} \tag{3.6}$$

$$P^+ = \frac{TP}{(TP + FP)} \tag{3.7}$$

Where TP is the true positive, FP the false positive and FN is the false negative.

3.3.1 Results of the LBNP dataset

The performance of the method on the ECG recordings of subjects undergoing the LBNP when detecting P, QRS, and T waves is depicted in Tables 3.2, 3.3, and 3.4, respectively. A total number of 173538 cycles from 90 cases are processed. Table 3.2 shows that the algorithm produces 99.77%

and 99.78% for sensitivity and positive predictability of P wave detection. Table 3.3 shows the sensitivity for QRS detection is 99.91% and the positive predictability is 99.86%. As in Table 3.4, the algorithm produces 99.74% sensitivity for T detection and 99.78% for positive predictability. The accuracies, are calculated as $100 - \sum i = 1nError_i$, for detecting P, QRS and T are 99.557%, 99.898% and 99.525%, respectively.

Table 3.2: Performance evaluation of the implemented ECG detection algorithm in detecting P wave for LBNP dataset

Stage	#Cycles	#TP	#FP	#FN	Error	Se (%)	P^+ (%)
1	32224	32192	34	32	0.2	99.90	99.89
2	27407	27388	29	19	0.175	99.93	99.89
3	29710	29681	15	29	0.148	99.90	99.94
4	32531	32503	23	28	0.157	99.91	99.93
5	27872	27831	42	41	0.297	99.85	99.84
6	15711	15692	24	19	0.273	99.88	99.85
7	6501	6481	19	20	0.599	99.69	99.71
8	1582	1568	13	14	1.7	99.12	99.18
Average					.443	99.77	99.78

Table 3.3: Performance evaluation for the implemented ECG detection algorithm in detecting QRS-complex wave for LBNP dataset

Stage	#Cycles	#TP	#FP	#FN	Error	Se (%)	P^+ (%)
1	32224	32206	17	18	0.1	99.95	99.94
2	27407	27391	19	16	0.12	99.94	99.93
3	29710	29695	13	15	0.09	99.95	99.95
4	32531	32514	23	17	0.12	99.94	99.92
5	27872	27855	16	17	0.11	99.93	99.94
6	15711	15698	14	13	0.17	99.91	99.91
7	6501	6490	16	11	0.04	99.83	99.75
8	1582	1579	8	3	0.07	99.81	99.5
Average					.102	99.91	99.86

Table 3.4: Performance evaluation for the implemented ECG detection algorithm in detecting T wave for LBNP dataset

Stage	#Cycles	#TP	#FP	#FN	Error	Se (%)	P^+ (%)
1	32224	32190	34	34	0.2	99.89	99.89
2	27407	27386	29	21	.18	99.92	99.89
3	29710	29680	15	30	.15	99.89	99.95
4	32531	32500	23	31	.16	99.9	99.93
5	27872	27829	42	43	.3	99.85	99.84
6	15711	15694	24	17	.26	99.89	99.85
7	6501	6477	19	24	.66	99.63	99.71
8	1582	1565	13	17	1.89	98.93	99.17
Average					0.475	99.74	99.78

3.3.2 Results of the MIT-BIH dataset

The detection performance for P, QRS complex, and T waves on MIT-BIH arrhythmia database are given in Tables 3.5, 3.6, and 3.7 respectively. A total number of 101579 cycles from 45 cases are processed. Table 3.5 shows that the algorithm produces 99.89% and 99.93% for sensitivity and positive predictability of detecting P wave. Table 3.6 shows the sensitivity for QRS detection is 99.96% and the positive predictability is 99.97%. In Table 3.7, the algorithm produces 99.81% sensitivity for T detection and 99.80% for positive predictability. The accuracies, calculated as $100 - mean(Error)$, for detecting P, QRS and T are 99.823%, 99.93% and 99.62%, respectively.

Table 3.5: Results of performance evaluation for the implemented ECG detection algorithm in detecting P wave for MIT-BIH dataset

Record	#Cycles	#TP	#FP	#FN	Error	Se (%)	P^+ (%)
100	2273	2270	5	3	0.35	99.8680	99.7802
101	1865	1862	1	3	0.21	99.8391	99.9463
102	2187	2185	2	2	0.18	99.9086	99.9086
103	2084	2083	3	1	0.19	99.9520	99.8562
104	2230	2226	3	4	0.31	99.8206	99.8654
105	2572	2567	6	5	0.43	99.8056	99.7668
106	2027	2022	2	5	0.35	99.7533	99.9012
107	2137	2135	0	2	0.09	99.9064	100
108	1763	1755	9	8	0.96	99.5462	99.4898
109	2532	2530	2	2	0.16	99.9210	99.9210
111	2124	2123	1	1	0.09	99.9529	99.9529
112	2539	2537	4	2	0.24	99.9212	99.8426
113	1795	1793	2	2	0.22	99.8886	99.8886
114	1879	1878	1	1	0.11	99.9468	99.9468
115	1953	1952	3	1	0.20	99.9488	99.8465
116	2412	2411	0	1	0.04	99.9585	100
117	1535	1532	1	3	0.26	99.8046	99.9348
118	2275	2274	2	1	0.13	99.9560	99.9121
119	1987	1985	0	2	0.10	99.8993	100
121	1863	1862	1	1	0.11	99.9463	99.9463
122	2476	2474	0	2	0.08	99.9192	100
123	1518	1517	2	1	0.20	99.9341	99.8683
124	1619	1617	1	2	0.19	99.8765	99.9382
200	2601	2600	0	1	0.04	99.9616	100
201	1963	1957	2	6	0.41	99.6943	99.8979
202	2136	2134	1	2	0.14	99.9064	99.9532
203	2982	2973	1	9	0.34	99.6982	99.9664
205	2656	2654	1	2	0.11	99.9247	99.9623
207	1862	1861	2	1	0.16	99.9463	99.8926
208	2956	2956	1	0	0.03	100	99.9662
209	3004	3002	1	2	0.10	99.9334	99.9667
210	2647	2645	2	2	0.15	99.9244	99.9244
212	2748	2745	2	3	0.18	99.8908	99.9272
213	3251	3250	1	1	0.06	99.9692	99.9692
217	2208	2206	1	2	0.14	99.9094	99.9547
219	2154	2153	0	1	0.05	99.9536	100
220	2048	2047	1	1	0.10	99.9512	99.9512
221	2427	2426	0	1	0.04	99.9588	100
222	2484	2482	0	2	0.08	99.9195	100
228	2053	2050	0	3	0.15	99.8539	100
230	2256	2254	1	2	0.13	99.9113	99.9557
231	1886	1884	0	2	0.11	99.8940	100
232	1780	1779	0	1	0.06	99.9438	100
233	3079	3077	0	2	0.06	99.9350	100
234	2753	2751	1	2	0.11	99.9274	99.9637
Average					0.177	99.89	99.93

Table 3.6: Results of performance evaluation for the implemented ECG detection algorithm in detecting QRS for MIT-BIH dataset

Type	#Cycles	#TP	#FP	#FN	Error	Se (%)	P^+ (%)
100	2273	2273	0	0	0.00	100	100
101	1865	1865	0	0	0.00	100	100
102	2187	2187	0	0	0.00	100	100
103	2084	2084	0	0	0.00	100	100
104	2230	2228	1	2	0.13	99.91	99.96
105	2572	2570	5	2	0.27	99.92	99.81
106	2027	2025	2	2	0.20	99.90	99.90
107	2137	2137	0	0	0.00	100	100
108	1763	1757	3	6	0.51	99.66	99.83
109	2532	2532	0	0	0.00	100	100
111	2124	2124	0	0	0.00	100	100
112	2539	2539	3	0	0.12	100	99.88
113	1795	1795	3	0	0.17	100	99.83
114	1879	1879	3	0	0.16	100	99.84
115	1953	1953	0	0	0.00	100	100
116	2412	2411	0	1	0.04	99.96	100
117	1535	1535	0	0	0.00	100	100
118	2275	2275	1	0	0.04	100	99.96
119	1987	1987	0	0	0.00	100	100
121	1863	1861	2	2	0.21	99.89	99.89
122	2476	2476	0	0	0.00	100	100
123	1518	1518	0	0	0.00	100	100
124	1619	1619	0	0	0.00	100	100
200	2601	2600	0	1	0.04	99.96	100
201	1963	1960	1	3	0.20	99.85	99.95
202	2136	2134	0	2	0.09	99.91	100
203	2982	2975	0	7	0.23	99.77	100
205	2656	2656	0	0	0.00	100	100
207	1862	1861	0	1	0.05	99.95	100
208	2956	2955	0	1	0.03	99.97	100
209	3004	3004	0	0	0.00	100	100
210	2647	2647	2	0	0.08	100	99.92
212	2748	2748	0	0	0.00	100	100
213	3251	3251	1	0	0.03	100	99.97
217	2208	2207	1	1	0.09	99.95	99.95
219	2154	2154	0	0	0.00	100	100
220	2048	2048	0	0	0.00	100	100
221	2427	2425	0	2	0.08	99.92	100
222	2484	2480	1	4	0.20	99.84	99.96
228	2053	2052	0	1	0.05	99.95	100
230	2256	2255	2	1	0.13	99.96	99.91
231	1886	1886	0	0	0.00	100	100
232	1780	1780	0	0	0.00	100	100
233	3079	3079	0	0	0.00	100	100
234	2753	2753	0	0	0.00	100	100
Average					0.07	99.96	99.97

Table 3.7: Results of performance evaluation for the implemented ECG detection algorithm in detecting T wave for MIT-BIH dataset

Type	#Cycles	#TP	#FP	#FN	Error	Se (%)	P^+ (%)
100	2273	2270	4	3	0.31	99.87	99.82
101	1865	1863	3	2	0.27	99.89	99.84
102	2187	2183	1	4	0.23	99.82	99.95
103	2084	2083	4	1	0.24	99.95	99.81
104	2230	2225	5	5	0.45	99.78	99.78
105	2572	2567	2	5	0.27	99.81	99.92
106	2027	2024	4	3	0.35	99.85	99.80
107	2137	2132	3	5	0.37	99.77	99.86
108	1763	1757	4	6	0.57	99.66	99.77
109	2532	2527	3	5	0.32	99.80	99.88
111	2124	2120	3	4	0.33	99.81	99.86
112	2539	2533	7	6	0.51	99.76	99.72
113	1795	1791	7	4	0.61	99.78	99.61
114	1879	1872	8	7	0.80	99.63	99.57
115	1953	1947	2	6	0.41	99.69	99.90
116	2412	2410	7	2	0.37	99.92	99.71
117	1535	1533	3	2	0.33	99.87	99.80
118	2275	2268	7	7	0.62	99.69	99.69
119	1987	1983	8	4	0.60	99.80	99.60
121	1863	1858	7	5	0.64	99.73	99.62
122	2476	2467	8	9	0.69	99.64	99.68
123	1518	1513	2	5	0.46	99.67	99.87
124	1619	1617	5	2	0.43	99.88	99.69
200	2601	2594	5	7	0.46	99.73	99.81
201	1963	1956	6	7	0.66	99.64	99.69
202	2136	2133	2	3	0.23	99.86	99.91
203	2982	2976	7	6	0.44	99.80	99.77
205	2656	2654	6	2	0.30	99.92	99.77
207	1862	1859	5	3	0.43	99.84	99.73
208	2956	2954	3	2	0.17	99.93	99.90
209	3004	2999	1	5	0.20	99.83	99.97
210	2647	2642	1	5	0.23	99.81	99.96
212	2748	2744	1	4	0.18	99.85	99.96
213	3251	3249	2	2	0.12	99.94	99.94
217	2208	2205	3	3	0.27	99.86	99.86
219	2154	2153	2	1	0.14	99.95	99.91
220	2048	2046	4	2	0.29	99.90	99.80
221	2427	2424	5	3	0.33	99.88	99.79
222	2484	2480	8	4	0.48	99.84	99.68
228	2053	2051	7	2	0.44	99.90	99.66
230	2256	2249	5	7	0.53	99.69	99.78
231	1886	1884	3	2	0.27	99.89	99.84
232	1780	1776	5	4	0.51	99.78	99.72
233	3079	3074	2	5	0.23	99.84	99.93
234	2753	2749	1	4	0.18	99.85	99.96
Average					0.38	99.81	99.80

3.4 Comparisons

As mentioned before, the literature focused on the detection of QRS complex only since it is easy to detect due to it's high amplitude comparable to the surrounding waves. Therefore, the comparisons in this section are between the accuracy of detecting QRS using the algorithm described in this chapter and other widely used methods in literature. The comparison between QRS detection using the implemented algorithm and other methods is presented in Table 3.8.

The results show that the method implemented in this dissertation performs better than other methods from literature.

Table 3.8: QRS detection comparison between the implemented QRS detection Algorithm and other important methods over the MIT-BIH dataset

Method	#FP	#FN	Error	Se (%)	P^+ (%)
ECG DT-CWT (this work)	31	39	0.07	99.96	99.97
Martinez et al. [83]	153	220	0.34	99.8	99.86
Pan et al. [99]	507	277	0.71	99.75	99.54
Li et al. [75]	65	112	0.17	99.89	99.94
Zhang et al. [134]	204	213	0.38	99.81	99.80
Ghaffari et al. [43]	160	213	213	99.80	99.85

3.5 Summary

This chapter presents a novel method for the detection of ECG deflection points based on a combination of Dual-Tree Complex Wavelet Transform (DT-CWT) and wavelet transform. For QRS complex, DT-CWT as the mother wavelet at level 4 is used. For P and T waves, db2 at level 2 is incorporated. Two datasets are used for reporting the results of the algorithm. The LBNP dataset contains 90 excerpts with more than 170,000 heartbeats. The MIT-BIH arrhythmia database con-

tains 45 cases with more than 100,000 cycles.

CHAPTER 4 Detection and Classification of Arrhythmia Severity

This chapter focuses on the creation of a computer-assisted decision making system for arrhythmia severity detection using machine learning algorithm and deterministic finite automate (DFA). Decision-making rules are based on the perceived dangers of certain ectopy and supported by the medical literature. Time-domain features as well as frequency domain features from Discrete Wavelet Transform (DWT) are extracted. The classification task is done using support vector machine method. Then, based on the rules, the system analyzes the vector using a deterministic finite automaton (DFA) to report severe abnormalities.

The obtained results of this system provide insights into the detection and classification of severe arrhythmia that may indicate cause for concern by health care providers that more serious cardiac event is occurring. In addition, it can be applied to applications where arrhythmia is an important factor to consider, such as blood volume loss systems.

The rest of this chapter is organized as follows. Section 4.1 briefly introduces the system. A description of the data is presented in Section 4.2. The method is described in Section 4.3 and the results are illustrated in Section 4.4. A comparison to other algorithms is presented in Section 4.5. The chapter is summarized in Section 4.6.

4.1 Introduction

Due to the potentially dangerous and life threatening nature of its complication, significant efforts have been devoted to develop automated arrhythmia detection systems. Any variation in the pacemaker sites or any delay in the conduction network of the heart may result in arrhythmias.

Even though imaging techniques, such as echocardiography or tallium scintigraphy offer detailed diagnostic evidences for some clinical applications, ECG plays a major role in the diagnosis of many cardiovascular diseases and malfunction, since it is non-invasive and low cost procedure.

Many studies have been conducted to explore effective ECG signal analysis and pattern recognition techniques for computer-aided diagnosis (CAD), such as arrhythmia. Arrhythmia can be categorized as life-threatening, such as ventricular fibrillation (VF) and tachycardia, and non-imminently life-threatening, such as ventricular ectopic beat (VEB) and supraventricular ectopic beat (SVEB). Although some types of arrhythmias are considered to have minimal consequences, if left untreated they may eventually lead to stroke or sudden cardiac death [104]. Hence it is prefeered that they are detected at any early stage.

Some abnormalities appear infrequently, such as sequences of heartbeats with abnormal timing or ECG wave form(s). For such abnormalities, often up to a week of ECG recordings recording by a Holter ECG monitor devices may be required to successfully capture the abnormalities [28]. Beat-by-beat analysis and classification of ECG can provide important information regarding the subject's cardiac condition [104]. Manual classification of heartbeats from long ECG recordings can be very time-consuming, therefore a reliable automated method to process, analyze, and classify arrhythmias from the hidden information encapsulated in patterns such as the intervals and amplitudes of ECG waves is of considerable importance.

The goal of this system described in this chapter is to go one step further than classifying the arrhythmia type recorded by the ECG; to automate the determination of the severity of arrhythmias detected by the ECG signal. The innovative system integrates several simple rules developed from the medical literature and formulate these rules in a deterministic finite automaton (DFA). This combination is novel to March of 2011, no similar systems have been developed.

In this system, features are extracted from time and frequency domains to train a model that classifies a heartbeat into one of three functional classes; normal (N), ventricular ectopic beat (VEB) or premature atrial contraction (PAC). Based on the classification, the output is checked according to a set of predefined rules formed from medical literature to report if the arrhythmia recorded in the signal is severe or not.

The algorithms used in this system are lucid, fast and amenable to real-time implementation. Figure 4.1 shows the overall process for the method. Next section elaborates on the details.

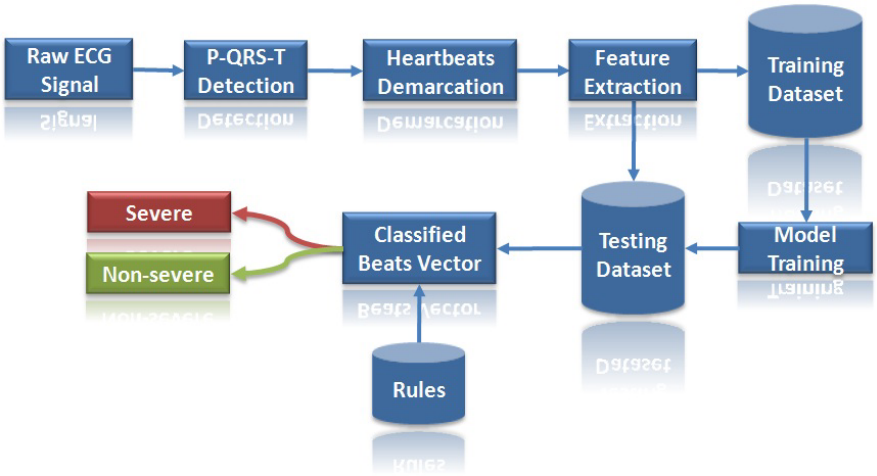


Figure 4.1: Schematic diagram of arrhythmia classification and severity detection system

4.2 Description of Dataset

The MIT/BIH arrhythmia database is a widely used database by the community to test and compare different methods [81]. Forty-eight excerpts, each contains two ECG lead signals with 11-bit resolution over a $10mV$ range that are band-pass filtered at $0.1Hz$ to $100Hz$ and sampled at $360Hz$.

Grouped into fifteen classes, the dataset has more than 100,000 labeled beats. Normal (N), premature atrial contraction (PAC) and ventricle ectopic beat (VEB) are the only classes considered in this study. The annotations of the other beat types are considered normal in this system, since

the rules do not relate them to severe arrhythmia. The grouping of different types is presented in Table 4.1.

Table 4.1: MIT/BIH mapping into three functional classes

Class Number	Short Name	#of Cycles	MIT/BIH Beats
1	N	91413	Normal, Left and Right Bundle Branch Block (LBBB and RBBB), Atrial escape beats (AE) Fusing of Ventricular, Normal beat (fVN) Paced beat (P), fusion of Paced Normal beat (fPN), and Unclassified beat (U)
2	PAC	2897	Atrial Premature beat (AP), and aberrated Atrial Premature beat (aAP)
3	VEB	7269	Premature Ventricular Contraction (PVC), SVEB and ventricular escape beat (VE)

4.3 Methodology

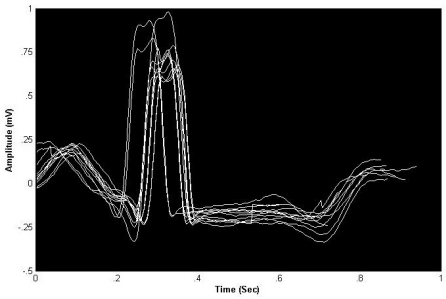
The system implemented in this dissertation is an automated one. Figure 4.1 presents the schematic diagram of the system.

Two groups of raw ECG signals are used; training and testing datasets. First, for each signal in both datasets, the characteristic points (P,QRS and T) of the heartbeats are identified by the detection algorithm introduced in Section 3.2. Then, the start and end of each beat are found using the information from the previous step. From each heartbeat, two types of features are extracted; time and Frequency domain features. The next step is to train a support vector machine (SVM) model using only the extracted features from the beats of the training dataset. The model is then used to classify the testing dataset beats into one of three functional classes (N, VEB, or PAC) as mapped in Table 4.1. The result is a vector of classified beats. Finally, an overlapping window of 30-last classified beats is analyzed according to several simple rules developed from the medical literature. After analyzing the vector. The results are provided.

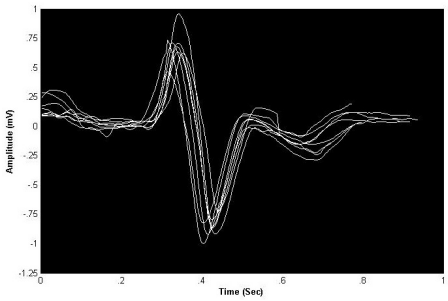
4.3.1 P-QRS-T Detection

For the purpose of detecting the ECG characteristic points, the method discussed in Chapter 3 is used. One of the properties that the detection method has is that it can detect whether R wave has positive or negative deflection. This can helps greatly in the discrimination between VEBs and other types of heartbeats when features are extracted from the information of the characteristic points, such as the amplitude of R, the ratio or T amplitude to R and the ratio of P amplitude to R.

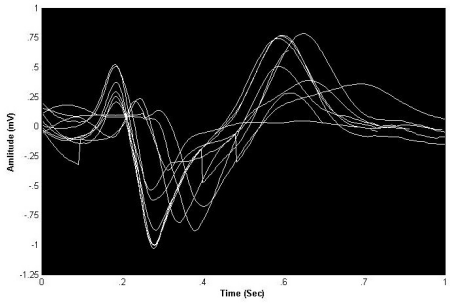
Figure 4.2 provides an insight about how important such features can be used as attribute to train a machine learning algorithm model that can discriminate between the different types considered in this dissertation.



(a) 10 normal heartbeats



(b) 10 PAC heartbeats



(c) 10 VEB heartbeats

Figure 4.2: A sample of 10 beats of normal (a), PAC (b) and VEB (c), as annotated in the MIT/BIH arrhythmia database

4.3.2 Heartbeats Demarcation

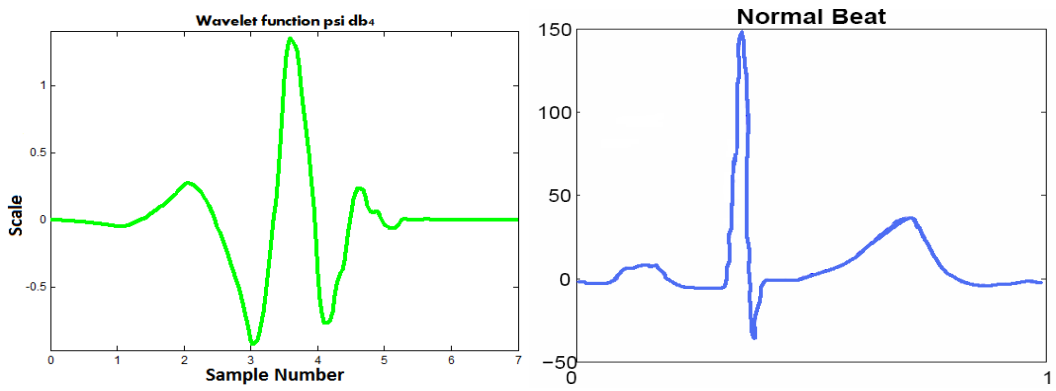
The start and end of each beat are found using the information from theP-QRS-T detection step. After detecting the deflection points (P, QRS and T), information from the P and T peaks is used to mark the beginning and ending of each heartbeat in the signal. The beats are then forwarded to a feature extraction step.

4.3.3 Feature Extraction

The observed pattern of beats in the ECG signal are often extremely complicated. Therefore, the features that are extracted must be directly related to the types the system tries to distinguish. For example, time-domain features such as ST interval, QT prolongation and QRS width indicate that the heart is functioning abnormally if the variations of these intervals are different than normal [56, 65, 111, 114].

As important as time-domain features, frequency-domain features can provide information about the abnormality of a heartbeat. For example, the complexity of a beat decreases when it is not normal [92]. In this system, frequency domain features are considered as well. To measure the complexity of a heartbeat, we apply wavelet transform to the beat. The beat is transformed using Daubechie 4 (db4) as the mother wavelet at level 2. The dominant shape in the ECG is a normal beat. As mentioned before, there is no standard rule to choose the mother wavelet. In this system, db4 is chosen because it resembles the normal beat (see Figure 4.3).

The transformation is followed by summing the detail coefficients that are close to zero. In addition, the sum of the absolute value of high frequency coefficients, the maximum and minimum, the median, just before median, and just after median coefficients are the features that are are computed and incorporated in this dissertation. A summary of features is provided in Table 4.2.



(a) Daubechie 4 (db4) (b) Normal neat

Figure 4.3: The mother wavelets used in this method

Table 4.2: The set of the features used for arrhythmia system

Number	Abbreviation	Full Name
1	QT	QT-interval
2	ST	ST-segment
3	QS	QRS width
4	PQ	PQ-interval
5	PR	PR-interval
6	HF_s	Sum of the Absolute Value of High Frequency Coefficients
7	Min_c	min Detailed Coefficient
8	Max_c	max Detailed Coefficient
9	Med	Medium of Detailed Coefficients
10	JAmed	Just After the Median
11	JBmed	Just Before the Median

Finally, the examples are divided into testing and training groups. As mentioned before, each signal in the dataset lasts for 30 minutes. The beats in the first five minutes (roughly 16%) are considered for training and the rest of the beats are taken as the testing dataset.

4.3.4 Model Training

Several machine learning methods have been dedicated for binary problems, in which there are only two output classes [26]. When the training data contains the labels, the problem is called

supervised learning, otherwise its called unsupervised. Another emerging type of learning is called semi-supervised learning; when some of the data have known class types [18].

Since linear methods are inefficient even in the classification of simple functions such as exclusive disjunction [58], non-linear algorithms for medical applications are widely used [77, 92].

One of the most important machine learning algorithms in the biomedical field is support vector machine (SVM), introduced by Cortes and Vapnik [25].

SVMs are based on the concept of decision hyperplanes that define decision boundaries between the different classes. The hyperplane separates a group of examples having different class labeled into positive and negative sets in n-dimensional space. A special nonlinear function called kernels are used to preserve the maximum distance between two classes into the transformed n-dimensional space [132]. We used SVM since it is widely used in medical applications [27, 66, 91].

The advantages of SVM over neural networks is in the cost function used. This means that SVM try to maximize the margin between the closest data points in different classes whereas NN tries to minimize the sum of square error. In other words, the effect of overlapping in generating a non-reliable model is decreased when the solution is provided by SVM.

SVM has disadvantages too. The main concern about using SVM is that they are computationally expensive when training a classification model. Moreover, to find the best kernel that can separate the data needs a huge effort. However, once SVM is used to train the model, testing of new data points is relatively fast. Another disadvantage for SVM is that it can generate an over fitted model if the classes in the dataset are imbalance. This problem is identified in this study and a method is provided to overcome the issue.

Given a set of N examples $S = (x_1, y_1), ..., (x_n, y_n)$, where $y_i \in \{+1, -1\}$, $x \in \mathfrak{R}^d$, and d is the dimensionality of the input. Training a model by SVM involves the minimization of the error

function:

$$\frac{1}{2}W^T W + C\sum_{i=1}^{N}\xi_i \tag{4.1}$$

subject to:

$$y_i\left(W^T\phi\left(x_i\right)+b\right) \geq 1-\xi_i \; and \; \xi_i \geq 0, \;\; i = 1, ..., N$$

where C is the capacity constant, W is the vector of coefficients, b a constant and ξ_i are parameters for handling input data. The kernel (ϕ) is a mapping from the independent vector X to the feature space. The most popular kernel function is the radial bases function (RBF) as provided in (4.2).

$$\phi = exp\left(-\gamma\left|x_i - x_j\right|^2\right) \tag{4.2}$$

The error function in (4.1) is minimized when $W = \sum_i \alpha_i\gamma_i\phi_i$, where ϕ_i is the kernel to be used for object i, γ_i is how wide the Gaussian function over example i, and α_i represent the influence data point i have on the decision function. Non-zero value of α_i makes object i a support vector (SV).

In this dissertation, a non-linear RBF-SVM model is trained using only the training examples obtained from the first five minutes as mentioned before. A publicly free software called LibSVM 2.89 is used to generate the model [12].

The values of the capacity constant C and γ of the kernel have direct influence on the model. For example, the larger the C, the more the error is penalized and γ is related to how wide the Gaussian is positioned around the data point. Thus, to abase over fitting, they should be chosen with care. It is also important to choose different capacity constant for different classes when the

dataset is imbalanced [1]. In this study, the best values of the capacity constant for each class and γ are obtained using a try and error method.

4.3.5 Vector Generation of Classified Beats

In this step, the trained model created before is used to classify the beats of the other 25 minutes of the signal (the testing dataset). This dataset consists of features generated from the heartbeats in ECG signal from five minutes and up to the end of that signal.

It is important to note that the model is trained using the first five minutes of all the 45 excerpts that the database has, and is tested one subject at a time incorporating the beats from after the five minutes to the end of that signal. This means that the overlapping exists between the heartbeats from the same signal is minimized. As a result, the constructed model is invulnerable to over fitting and hence have high reliability.

In order to classify a given unknown example X_k using the model created before, we simply calculate the following function:

$$f(X_k) = sgn(y_i < W, \phi(X_i) > -b) \tag{4.3}$$

After successively classifying every heartbeat in the testing dataset using the formula in (4.3), a vector that represent these classifications is generated and then analyzed by rules supported by the medical literature. Next section elaborates how the analysis is done.

4.3.6 Analysis of the Classified Beats with Deterministic Finite-State

It is important to point out that the crux of this system is to detect how severe the arrhythmia is. The result is a report of whether the signal contains severe abnormality or not. To generate such a report, three simple rules are formulated based on the medical literature that would give health

care providers information that a severe heart arrhythmia is occurring or is about to occur.

As Figure 4.4 shows, the rules are instituted as a deterministic finite automate (DFA) to analyze the vector of classified heartbeats generated before. A window of 30 classified beats is considered for the analysis. This is because the study is aimed to discover a severe abnormality as soon as possible and this window size converges with the rules. Moreover, the window size is believed to be clinically relevant.

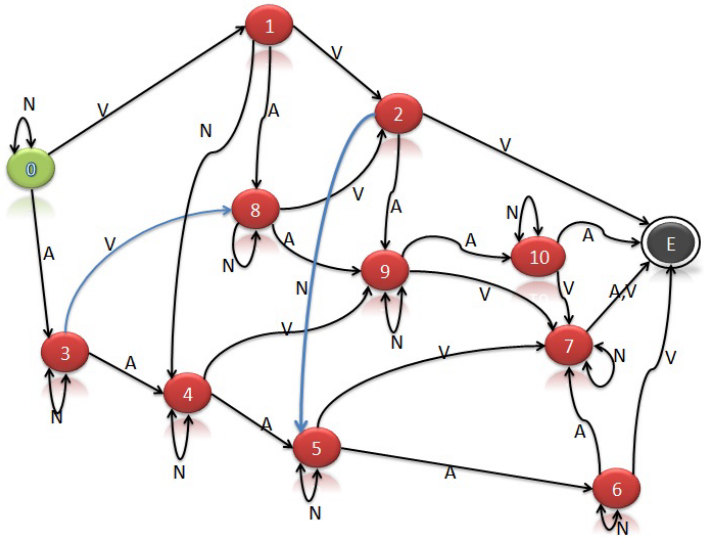


Figure 4.4: The deterministic finite automate (DFA) of the rules used in this study. State 0 represents the start state and *E* is the final (severe) state. The possible alphabets of the DFA are N, A, or V. Only one alphabet is received at a time and the input causes a transition to only one state

The analysis is done by a sliding window of five beats. The rules that are considered in this study are:

- **Rule 1:** Three consecutive VEBs in a window

  When analyzing the window and there are three VEB beats in a row, then the whole signal is considered as severe. Finding three VEB beats in window is a concern for the potential development of more lethal arrhythmias such as ventricular tachycardiac which may lead

to cardiovascular collapse and/or sudden cardiac death or be an indication of developing ischemia [67, 90].

- **Rule 2:** A minimum of six PAC beats in a window

  If the analysis resulted in finding six PAC beats (but not necessarily in a row), that signal is reported as severe. Six PAC beats represent 20% of the window. While PAC beats are benign in the healthy heart, they can herald the onset of more serious arrhythmias such as atrial fibrillation and atrial flutter in a diseased heart. [52, 55, 96].

- **Rule 3:** The following rules were created since the combination of VEBs and PACs may indicated cardiac irritability from a number of causes ranging from respiratory pathology, cardiac toxicity of medications to circulating electrolyte imbalances [49, 50, 64, 128].

  Rule 3.a: Four PAC beats adjoin with one or more VEB.

  Rule 3.b: Two consecutive VEBs with one or more PACs.

  Rule 3.c: Two VEBs not in a row with three or more PACs.

The mathematical model of the diagram in Figure 4.4 is given below:
A deterministic finite-state machine is a quintuple $(\Sigma, S, s_0, \delta, F)$ [89] where:

- Σ is non-empty set of alphabet defines the accepted input. The accepted input for this study DFA are {N, A, V}, where N means a normal beat in the classified vector, A is a PAC beat, and V is a VEB.

- S is a finite, non-empty set of states. As Figure 4.4 shows, the states for our DFA are {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, E}.

- s_0 represents the start state and must $\in S$. Clearly, the start state is 0 as shown in the figure.

- δ is the function used to transition from one state to another when one of the acceptable alphabets is read. Each transition returns only one state $\in S$. The mathematical notation of this function is $\delta : S \times \Sigma \rightarrow S$.

- F is a set of final states and it must $\in S$. In our diagram, the final state is E.

Figure 4.4 can also be represented as a state-transition table. Typically, state-transition tables are two-dimensional where the vertical dimension represents events and horizontal dimension is the current state. The state-transition table for Figure 4.4 is given in Table 4.3.

Table 4.3: The state transition table from the deterministic finite automate (DFS) of Figure 4.4. The start state is 0 and the end state E represents a severe arrhythmia

State Transition Table			
Event ⇒ State ⇓	**N**	**A**	**V**
0	0	3	1
1	4	8	2
2	5	9	E
3	3	4	8
4	4	5	9
5	5	6	7
6	6	7	8
7	7	E	E
8	8	9	2
9	9	10	7
10	10	E	7
E	E	E	E

In this system, when a vector is generated, the alphabets of the vector are considered as an expression and is shoved to the DFA. As the expression is evaluated, the system triggers a search to the final state E and if visited, a severe arrhythmia is reported. If the alphabets of the window

did not lead to a finial state, the window is shifted five beats and the procedure is applied again. If none of the windows that are checked resulted in a visit to the final state, the signal is reported as non-severe.

4.4 Results

In this section, the results for arrhythmia classification and severity detection are illustrated. All data processing was performed off-line using a commercial software package (MATLAB 7.7, The MathWorks Inc., Natick, MA, 2000). The classification of arrhythmia is done by support vector machine learning algorithm as implemented in LibSvm 2.89 [12]. A radial basis function is used as the kernel.

The number of beats in training and testing datasets are provided in Table 4.4 below. As it can be seen in the table, the datasets for training and testing are imbalanced. There is an imbalance of 13 to 1 between N and VEB classes and it reaches to 55 to 1 between N and PAC classes in the training dataset. This leads to the necessity to provide a different capacity constant C for each class when the model is trained [1].

Table 4.4: The number of beats for the functional classes (N, PAC and VEB) as extracted from the MIT/BIH database. The training beats are generated from the first five minutes of each signal. The rest of the beats are counted in the testing dataset. If the annotation in the database is not PAC or VEB then it is considered as normal

Class	**#Beats**	**#Training**	**#Testing**
N	91413	15674	75739
PAC	2897	282	2615
VEB	7269	1211	6058

The results of the classification are presented by Sensitivity (Se), Specificity (Sp) and accuracy as in equations (3.6), (4.4) and (4.5). True Positive (TP), False Negative (FN), True Negative (TN)

and False Positive (FP) are also computed and provided in the results.

$$Sp = \frac{TN}{(TN + FP)} \tag{4.4}$$

$$Accuracy = \frac{TP + TN}{(TP + +TN + FN + FP)} \tag{4.5}$$

4.4.1 Arrhythmia Classification Results

The classification model is first trained with the training dataset and then it is used to classify the testing dataset. Before training the model, the best values of γ and the capacity constants C for each class must be found with care since their values will affect the accuracy. To find these values several runs are performed and illustrated in Subsection 4.4.1.1. The results of the trained model is given in Subsection 4.4.1.2. Subsection 4.4.1.3 shows the results for the testing dataset.

4.4.1.1 Finding the capacity constants and γ to train the model

As shown in Table 4.4, the training dataset is imbalanced. So, to find the best values for γ and the capacity constants for each class, a trial and error method is performed. The g-means metric suggested by Kubat et al [69] is used to measure the performance of the trained model. The g-mean metric has been used by several researchers for evaluating classifiers on imbalanced datasets [1, 68, 69, 130]. Equation 4.6 shows the definition of this metric, where acc+ and acc- are the average sensitivity and specificity of the model, respectively.

$$g = \sqrt{acc + \dot{a}cc-} \tag{4.6}$$

In each run, one parameter is picked and increased in a step of multiple of 2. For example, the first run is performed with C_1, C_2, C_3 and γ are all = 2, where C_1, C_2 and C_3 stand for the capacity

constant for N, PAC and VEB, respectively. The second run with C_1=4 and the other parameters are equal to 2 and so on until the value is equal to 1024. This scenario resulted in 10,000 runs and in each execution the g-mean is calculated and stored along with the value for each parameter. The highest value of g-mean is then used to obtain the values of the parameters. The best values provided by the g-mean are found to be 1024, 64, 128 and 0.5 for C_1, C_2, C_3 and γ, respectively.

4.4.1.2 Training Model Results

The training examples are obtained from the first five minutes of each subject in the database. A total of more than 17,000 annotated beats are used. The value of each parameter from the previous subsection is used. The results of the trained model with 10-fold cross-validation are illustrated in Table 4.5. The accuracy of the model is 99.54%.

Table 4.5: Sensitivity and specificity of arrhythmia classification model using the training dataset with 10-fold cross-validation

Class	**#TP**	**#TN**	**#FP**	**#FN**	**Se** (%)	**Sp** (%)
N	15613	1445	32	61	99.61	97.83
PAC	269	16789	20	13	95.39	99.88
VEB	1176	15882	57	35	97.11	99.64

4.4.1.3 Testing Results

The trained model created before is used to classify each heartbeat in the testing dataset. As mentioned before, the testing dataset is created from the heartbeats in each signal from after five minutes up to the end of that signal. The results are provided in Table 4.6 below. The accuracy of the model is 93.52%.

Table 4.6: Sensitivity and specificity of arrhythmia classification model using the testing dataset

Class	**#TP**	**#TN**	**#FP**	**#FN**	**Se** (%)	**Sp** (%)
N	68923	7622	882	6816	91.00	89.63
PAC	2267	74278	1772	348	86.69	97.67
VEB	5355	71190	5213	703	88.39	93.18

4.4.2 Results for Detection of Arrhythmia Severity

The novelty of this system is to detect the presence of and classify the severity of arrhythmia according to several simple rules supported by the medical literature. In this study, the rules are implemented using a finite state machine. Table 4.3 provides the state transition table and Figure 4.4 shows the finite state automaton diagram. There are twelve states with 0 as the start state and the mnemonic *E* is the final (severe) state. The classes of beats in the vector generated before are read and inputted one by one to the DFA. When *E* state is visited, a severe arrhythmia is reported.

To substantiate the results, another analysis is done for the testing dataset without classification. The actual labels for the segmented beats are used. A vector is created for each signal from these labels and are fed to the DFA. This is because we want to create a gold standard to compare the results of the classified beats with. When analyzing the testing labels without classification, the DFA reports a severe arrhythmia in 29 cases and the other 16 subjects as non-severe.

According to the trained model, the system reports a severe arrhythmia in 25 cases. From the 45 subjects in the database, the system indicated a non-severe arrhythmia in 13 records producing an accuracy of the decision support system of 84.44%. In addition, we calculate TP, TN, FP and FN, sensitivity and specificity of the arrhythmia severity detection system as compared the the gold standard. Table 4.7 lists these values for the implemented system.

Table 4.7: Arrhythmia severity detection results. The golden measurement used for comparing is the analysis of the actual annotated beats of the testing dataset when provided as input to the DFA

Class	#TP	#TN	#FP	#FN	Se	Sp	Accuracy
Severe	25	13	3	4	86.21	81.25	84.44
NonSevere	13	25	4	3	81.25	86.21	84.44

4.5 Comparison

It is difficult to compare the result with other studies in literature for many reasons. First, different studies use different number of classes. Second, even though two studies have the same number of classes, the mapping for their classes might be different than the MIT/BIH annotations. Finally, some studies exclude some records from the database since they contain noisy signals.

To fairly compare the results between the algorithm implemented in this dissertation and other methods, the same mapping is used. In other words, the mapping used in this dissertation is altered to match the mapping that the other study uses. In addition, if the study excludes a specific record from the database, that record is also excluded when reporting the results from the implemented algorithm. Following this scheme, the results can fairly be compared.

The comparison between the implemented algorithm and other methods from literature is presented in Table 4.8. Some studies reported their results using sensitivity and specificity, whereas in some other studies only the accuracy is found. The features used in the created models are 11 as in Table 4.2. Each row in Table 4.8 shows the sensitivity (Se), specificity (Sp) and the accuracy of the created model using the map that the other study uses to categorize the annotated beats in MIT/BIH arrhythmia database.

Table 4.8: Comparison between arrhythmia classification model implemented in this dissertation and other method in literature

	This work			Other Study				
#Classes	Se	Sp	Acc	Se	Sp	Acc	#Features	Study
4	95.3	96.7	95.8	91	92	N/A	11	Exarchos et al. [37]
	95.3	96.7	95.8	97.78	97.78	N/A	19	Güler and Übeyli [45]
	95.3	96.7	95.8	N/A	N/A	93.46	19	Nasiri et al. [93]
	96.9	98.8	98.1	N/A	N/A	98.2	3	Tsipouras et al.[122]
	95.3	96.7	95.8	82.6	97.1	94	9	Hu et al. [51]
5	96.7	97.9	97.4	N/A	N/A	94.6	15	Chazal et al. [28]
	95.3	96.7	95.8	96.23	99.09	93.59	150	Kutlu and Kuntalp [72]
6	96.1	96.6	96.3	N/A	N/A	91.67	303	Melgani and Yakoub [87]
	96.1	96.6	96.3	N/A	N/A	97.33	11	Arifet et al. [3]
	97.3	98.7	98.5	99.6	95.1	98.9	17	Song et al. [116]
7	97.4	99.3	98.7	N/A	N/A	96.5	50	Lin et. al. [76]
	97.4	99.3	98.7	N/A	N/A	96	08	Osowski and Linh [97]

4.6 Summary

The novelty of this study is to detect the presence of and classify the severity of arrhythmia according to several simple rules developed from the medical literature using a decision support system and deterministic finite automaton. The system is divided into three sub-systems; ECG detection, arrhythmia classification and arrhythmia analysis. ECG detection sub-system is used to extract features form time-domain and wavelet transformation is applied to calculate frequency-domain features. An arrhythmia classification sub-system is implemented to train a model and use it to test the class of unknown beats into one of three function classes (normal, premature atrial contraction and ventricular ectopic beat). The classified beats are then used in the arrhythmia analysis sub-system to indicate ectopy, which may lead to heart dysfunction. The MIT/BIH arrhythmia database is used for this purpose.

The method described allows for the development of a computationally inexpensive method to

rapidly detect and report ectopy in an early warning manner that may allow health care providers for more time to react prior to possible patient deterioration. In this dissertation, the system implemented in this chapter can be used to extract informative features that tell something about the arrhythmia severity to predict the severity of blood volume loss as discussed in the next chapter.

CHAPTER 5 Loss of Blood Volume Prediction

In this chapter, a novel method for the severity prediction of loss of blood volume is introduced. The method extracts features from physiological signals and combines them with other features extracted from the arrhythmia severity detection system discussed in the previous chapter. Three biomedical signals are used in the analysis, namely ECG, ABP and impedance.

The rest of this chapter is organized as follow: Section 5.1 briefly introduces the approach. The method is illustrated in Section 5.2 and the results are given in Section 5.3. A comparison to other algorithms is presented in Section 5.4. The chapter is then summarized in Section 5.5.

5.1 Introduction

Hemorrhage can be a fatal consequence of a traumatic injury in both military and civilian settings. From physiological point of view, blood loss will influence many biomedical signals including, but not limited to, ECG, ABP, skin temperature, phonocardiogram (PCG), carotid pulse (CP) and impedance. Focusing on only one of such signals to build a model that is capable to distinguish the severity of hemorrhage means that some information may be ignored. However, analyzing all of them is time consuming and would substantially increase the complexity of the model. Therefore, a trade off should be applied when choosing the signals to be processed in order to detect the severity of blood loss.

Some questions arise here; what are the most informative signals? How many signals should be analyzed and processed and how easy they are acquired? The former two questions are implicit in the military setting. However, in civilian setting the answer to the first question is the most

important one, since a wider range of monitoring devices can be used. Nevertheless, the focus of this project is to create a system with the following properties:

1. The system should be capable of discriminating between severe and non-life-threatening hemorrhage; inevitably, achieving high accuracy.

2. The system should continuously provides feedback about the state of the subject. For example, at each point in time the feedback/recommendations is formed using the last few seconds/minutes of the collected signals.

3. The system should be efficient in terms of memory requirement and execution time. Moreover, it must be close to real-time.

Although, the second and third properties are very important, the first property seems to be the most important one, at least from medical point of view.

In this dissertation, a novel method for the prediction of the severity of blood volume loss is illustrated. The schematic diagram for the method is shown in Figure 5.1.

The method includes three main steps: Signal pre-processing, Feature Extraction, and Classification. The signals that are chosen for analysis are ECG, ABP and thoracic impedance. The dataset used is obtained from lower body negative pressure (LBNP) protocol. LBNP was proved to be a useful tool for simulating the early phase of acute hemorrhage in humans [22–24].

5.2 Methodology

As shown in Figure 5.1, the inputs to this system are from ABP, impedance, ECG detection algorithm and arrhythmia classification and severity detection system. Features from ABP, impedance

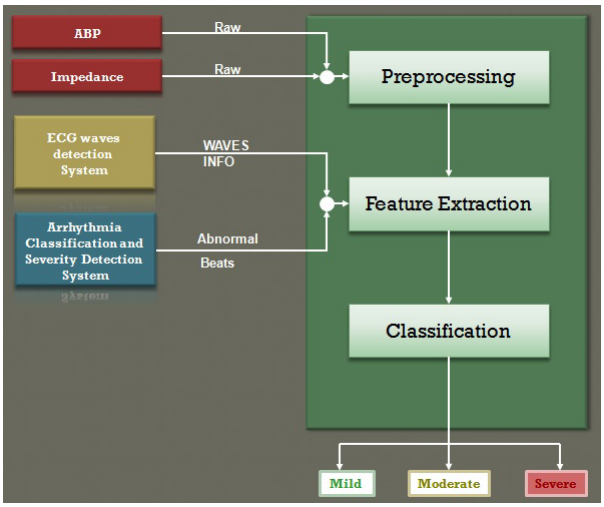


Figure 5.1: Schematic diagram for prediction of loss of blood volume severity system

and ECG detection algorithm and arrhythmia classification and severity detection system are extracted and used as the attributes for classifying blood volume loss.

The feasibility of analyzing multiple physiological signals to predict the severity of loss of blood volume is elaborated by creating and comparing five types of models, in which features extracted from, respectively:

1. ECG signal only.

2. ECG and ABP signal.

3. ECG, ABP and impedance.

4. ABP and impedance.

5. ECG and impedance

5.2.1 Description of the dataset

The LBNP dataset that is used in this work is described in Section 3.2.1.1. In this dataset different cases have different collapse stages; for example, some may collapse at stage 4, where someone else may collapse at stage 7. This observation calls for further attention when forming a prediction system. One solution to this problem is the division of the stages into some predefined categories identified by clinicians. This idea is adopted in this dissertation. The categorization of different collapse stages into classes is done by physicians. Table 5.1 presents the mapping of collapse stages into three classes (mild, moderate and severe). The main idea in forming this mapping is that the first stages (lower stages) are mapped to mild, the last few stages are mapped to severe and all middle LBNP stages are mapped to moderate.

Table 5.1: Mapping of different collapse stages into three class labels

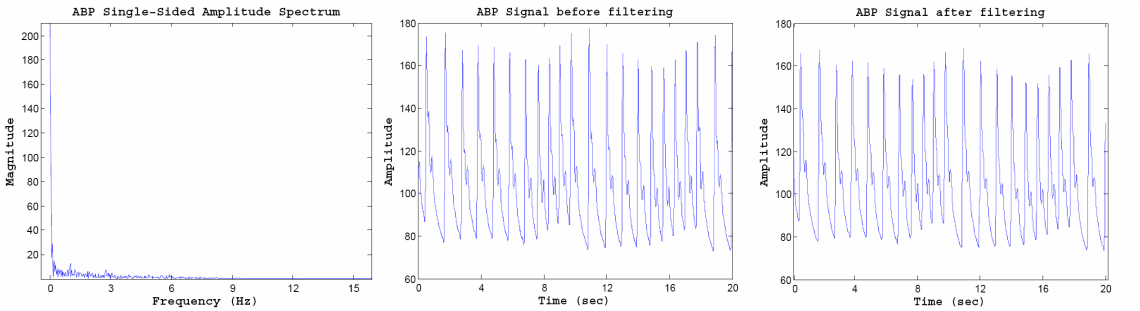
Collapse Stages	Mild	Moderate	Severe
4	Baseline and Stage 1	Stage 2	Stage 3 and 4
5	Baseline and Stage 1	Stage 2 and 3	Stage 4 and 5
6	Baseline, Stage 1 and 2	Stage 3 and 4	Stage 5 and 6
7	Baseline, Stage 1 and 2	Stage 3 and 4	Stage 5, 6 and 7

5.2.2 Signal Preprocessing

Several types of noise might exist in ABP and impedance signals. Therefore a suitable filter to enhance signal-to-noise ratio must be designed and used. For this purpose, a Butterworth bandpass filter is designed with stop and pass bands specified for each signal by applying Fourier transformation (FT) to identify the cut-off of the filter.

Figure 5.2a shows the magnitude of the Fourier Transform for ABP signal vs. frequency. Observing this figure and the knowledge of the signal, a low-pass filter that filters out frequencies

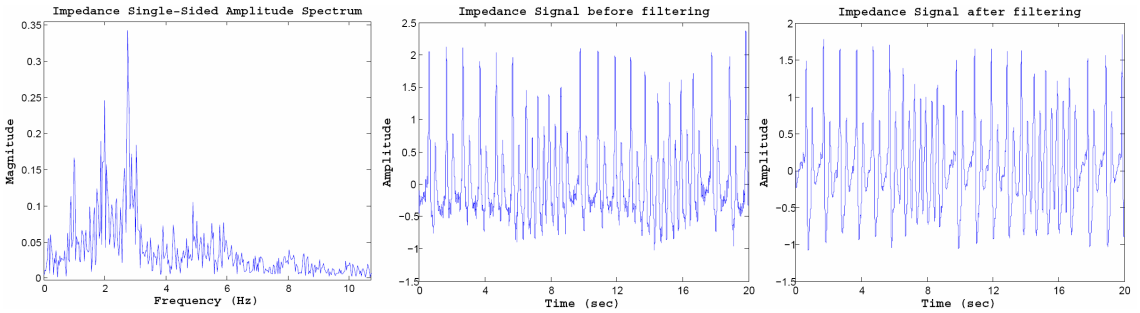
higher than $6Hz$ is designed for the ABP signal. It is important to mention that Figure 5.2a is not to scale in frequency. The results before and after applying the designed filter based on Fourier transform are provided in Figure 5.2b and Figure 5.2c, respectively.



(a) Fourier transform (b) ABP before filtering (c) ABP after filtering

Figure 5.2: The process of filter design for ABP signal

To design a filter for impedance signal, Fourier transform is also applied to the signal. Figure 5.3a shows the transformation. From the figure, a band-pass filter with cut-off of $0.8Hz$ and $10Hz$ is designed. Figure 5.3a is not to scale. The results before and after applying the filter are provided in Figure 5.3b and Figure 5.3c, respectively.



(a) Fourier transform of impedance signal (b) Impedance signal before filtering (c) Impedance signal after filtering

Figure 5.3: The process of filter design for impedance signal

5.2.3 Feature Extraction

A quick look at Figure 5.1 shows that the inputs for this stage are the filtered version of ABP and impedance signals, ECG characteristic points (i.e. P-QRS-T), and abnormal heart beats.

Two different sets of features are extracted from ECG signal; features from time domain and wavelet domain. The features are extracted from a non-overlapping window of size 20 seconds. This is because the study is aimed to discover a severe loss of blood as soon as possible. Moreover, the window size is believed to be clinically relevant. In the following subsections, the extracted features are illustrated.

5.2.3.1 Time Domain Features

1. ECG Intervals and Amplitude Ratios.

   The time domain features that are computed from ECG signal are the average intervals of the beats in the window. These intervals are PQ, QRS, PR, ST and QT. In addition, the standard deviation and median is computed for each interval and included. Moreover, the average ratio of amplitudes are also calculated. These ratios are P-to-Q, Q-to-S, P-to-R, S-to-T and Q-to-T.

   There are 20 features extracted from the ECG characteristic points.

2. Statistical Measurements.

   The mean, standard deviation and median are the statistical properties that are considered in this dissertation for each signal (ABP, ECG and Impedance). In addition, Heart rate variability (HRV) signal is calculated from the ECG and the aforementioned measurements are computed from the obtained signal. RRI is proved medically to be a good feature that

indicate blood loss when used in such a system. However, it is not included in the feature set because the mean of HRV signal is already considered.

Thus 12 statistical features are extracted from the signals.

3. Signal Complexity and Mobility.

For each signal considered in this dissertation, the complexity and mobility are calculated and included in the feature set. These two features are computed since they can measure the level of variations among any signal quantitatively and are often used in biomedical applications to quantify the first and second order variations of the signal [92].

Assume a signal x_i, $i = 1, ..., N.$, vector D represents the first order variations in X, and vector G represents the second order variations such that:

$$d_i = x_i - x_{i-1}$$

$$g_i = d_i - d_{i-1}$$

Now, using the signal vector X and vectors D and G, the following first and second order factors are defined:

$$S_0 = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}$$

$$S_1 = \sqrt{\frac{\sum_{i=2}^{N-1} d_i^2}{N-1}}$$

$$S_2 = \sqrt{\frac{\sum_{i=3}^{N-2} g_i^2}{N-2}}$$

Then, the complexity and mobility are defined as equations (5.1) and (5.2), respectively.

$$Signal\ Complexity = \sqrt{\frac{S_2^2}{S_1^2} - \frac{S_1^2}{S_0^2}} \tag{5.1}$$

$$Signal\ Mobility = \frac{S_1}{S_0} \tag{5.2}$$

4. KullbackLeibler Distance. Given two probability distributions P and Q, Kullback-Leibler Distance (KLD) is a non-symmetric measure of the divergence between those two distributions [70]. Typically P represents the ”true” distribution and Q represents the approximation of P.

   For probability distributions P and Q of a discrete random variable their KLD is calculated from equation (5.3).

$$KLD(P \parallel Q) = \sum_{i=1}^{N} P(i) log \frac{P(i)}{Q(i)} \tag{5.3}$$

The ECG, ABP and Impedance are discrete stochastic signals and therefore form discrete random variables. In this work, three subjects are randomly selected and for their signals, the probability distribution function is calculated. That means, there are 9 functions; three for ECG, three for ABP and three for Impedance.

Let P^i_{ECG}, P^i_{ABP} and P^i_{Imp} define the probability distribution functions of the i^{th} randomly selected subject. Now, to extract features form each subject in the dataset, KLD, as defined in equation (5.3), is calculated between each of the predefined functions and the probability distribution functions computed from the subject's signals. That means, nine features are extracted from each subject. These features are $KLD(P^i_{ECG} \,||\, Q_{ECG})$, $KLD(P^i_{ABP} \,||\, Q_{ABP})$ and $KLD(P^i_{Imp} \,||\, Q_{Imp})$, where Q_{ECG}, Q_{ABP} and Q_{Imp} represent the probability distribution functions obtained from the subject's ECG, ABP and Impedance, respectively.

5. Rate of Abnormal Heartbeats

The arrhythmia classification and severity detection system that was discussed in Chapter 4 is used to extract the rates of abnormal beats in the window defined earlier. The number of PACs and VEBs are the two features considered from the arrhythmia system. The novelty of using such features is in quantitatively exploring the relationship between arrhythmia and blood loss. This study further explores the observation that the rate of arrhythmic beats increases as a result of hypoxia, whereas hypoxia becomes very likely during an acute blood loss. The relationship between hypoxia and arrhythmia is explained on the metabolic basis [19, 71, 101, 129].

The prediction system feeds the whole window to the arrhythmia system. The result will be a vector of classified beats. From this vector, the number of VEBs and PAC are counted and

included.

The actual arrhythmia system used here is slightly different from the one discussed in Chapter 4. The system used here is trained with the whole ECG signal instead of the first five minutes. This is because the model is applied to a different database. This means that the model was trained with the MIT-BIH beats, and then the trained model was used to classify the beats in the LBNP database. It it important to note that the two signals are not the same. To standardize both datasets, a normalization in terms of amplitude is applied to be between $[-1, 1]$. The extracted features are also normalized by the sampling rate, if and only if, the feature is generated from time domain. For example, PQ is used as a feature and it is extracted from time-domain. As such, this feature is normalized by the sampling rate to standardize the concept of this feature in both datasets.

It was observed that the rate of abnormal beats due to hypoxia increased substantially in the case of acute blood loss, which would make such features very important to distinguish severe hemorrhages.

The total number of time-domain features extracted, when the ECG, ABP, and Impedance signals are considered is 49 attributes. That is, 33 features from the ECG and 8 from each of the ABP and impedance signals.

5.2.3.2 Wavelet Domain Features

To extract features from the frequency domain, wavelet transform is applied with daubechie 4 (db4) at level 4. The resulting detail coefficients of levels 1 to 4 are used to calculate statistical features. For each level, the standard deviation, mean, median, kurtosis, and skewness are computed. The approximation coefficients at level 4 are used, as well as the detail coefficients to

calculate the entropy of the frequency domain as described in equation (5.4).

$$Entropy = -plog(p), \tag{5.4}$$

where p is calculated as follows:

$$p = \frac{\sum_{j=1}^{N} |ac_j|^2}{N \sum_{l=1}^{4} \sum_{i=1}^{M} dc_i}$$

where acj's are the approximation coefficients at level 4, N is the number of these coefficients and dci's are the detail coefficients at level l.

Empirically, the best results are achieved when db4 at level 4 is used for the transformation. The total number of features extracted from the frequency domain is 21 attributes form each incorporated signal.

5.2.4 Classification

The machine learning algorithm, SVM, introduced in Section 4.3.4 is applied to predict the severity of loss of blood volume. As mentioned before, five models are created in order to emphasis the importance of incorporating multiple physiological signals. The labels for the examples is mild, moderate or severe.

The RBF kernel function in (4.2) is adopted for all models. The SVM parameters for each model are selected based on an optimization tool installed from LibSvm [12]. The values for the parameters for each model are given in the following section.

5.3 Results

In this study, 90 subjects from the LBNP dataset are used. For each stage, a non-normalized and non-overlapping window of size 20 seconds (10000 samples) is defined. Three types of signals (ECG, ABP and impedance) are used to extract a total of 112 features. The labels for the examples

are mild, moderate or severe as defined in Table 5.1. The sensitivity and specificity are calculated using equations (3.6) and (4.4), respectively. The accuracy is calculated using equation (5.5). For model validation, two types of validations are considered; 10-fold cross validation and leave one subject out cross validation. The results are illustrated in Table 5.2.

Leave one subject out cross validation is computed as training the model with $n - 1$ subjects ($n = 90$), then the trained model is used to test the excluded subject. From each subject, several examples are obtained using a non-overlapping sliding window of size 20 seconds. This means that all the examples generated from a subject will be tested by a model that was trained with the examples obtained from the $n - 1$ subjects. This way will decrease the dependency between the examples and a more reliable model is created. Moreover, this way simulates the model in real medical application.

$$Accuracy = 1 - \frac{\#missclassified}{Total\ number\ of\ examples} \tag{5.5}$$

Time complexity is measured using an eight cores computer, where each is $1.67GHz$. The size of the model is given as the space required to store a serialized JAVA object from the model. The capacity constant C and σ are obtained from a tool installed with LibSvm 2.89.

Figures 5.4 and 5.5 visually illustrate the results in Table 5.2 using 10-folds cross validation and leave one subject out cross validation, respectively.

Table 5.2: Results for the different blood loss models that are created in this dissertation. The classes are mild (2670 examples), moderate (1745 examples) and severe (2052 examples)

		Signals Used				
		ECG Only	ECG & ABP	ECG, ABP & IMP	ABP & IMP	ECG & IMP
Parameters	C	2^5	2^5	2^7	2^5	2^5
	σ	2^{-3}	2^{-5}	2^{-5}	2^{-1}	2^{-1}
	#Features	54	83	112	58	83
	Model Size	2375 KB	3834 KB	4302 KB	2926 KB	3941 KB
	Time	26.13sec	36.56 sec	40.74 sec	34.66 sec	32.73 sec
10 Folds CV	Se	89.4	90.7	91.6	82.3	91.4
	Sp	94.7	95.2	95.6	91.2	95.7
	Missclassified	686	604	543	1146	555
	Acc	89.39	90.66	91.6	82.28	91.42
Leave one subject out	Se	73.8	74.6	79.6	60.1	74.1
	Sp	78.1	79	85.3	68.4	77.9
	Acc	75.95	76.8	82.45	64.2	75.7

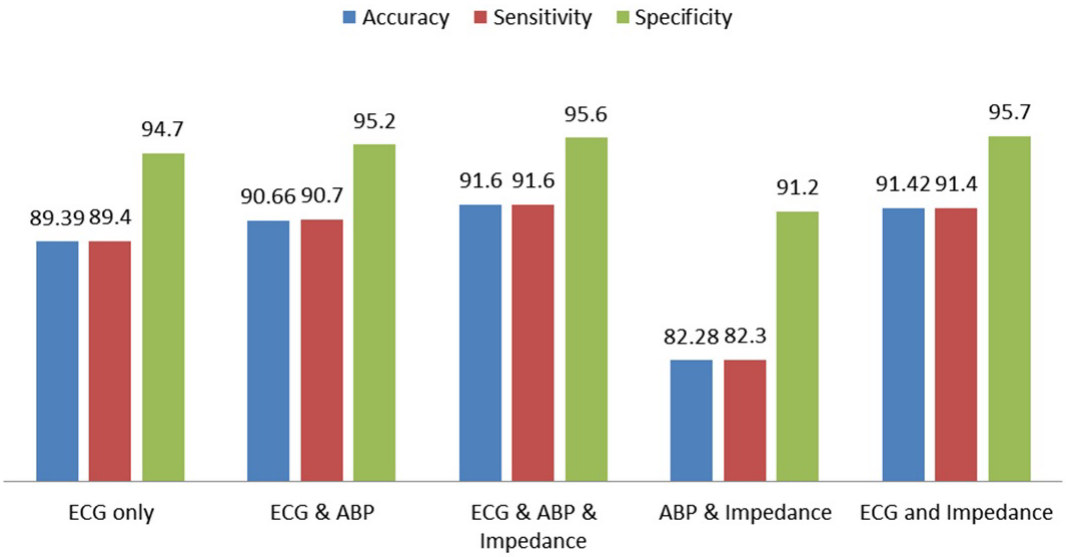


Figure 5.4: Results of the created models using 10 folds cross validation

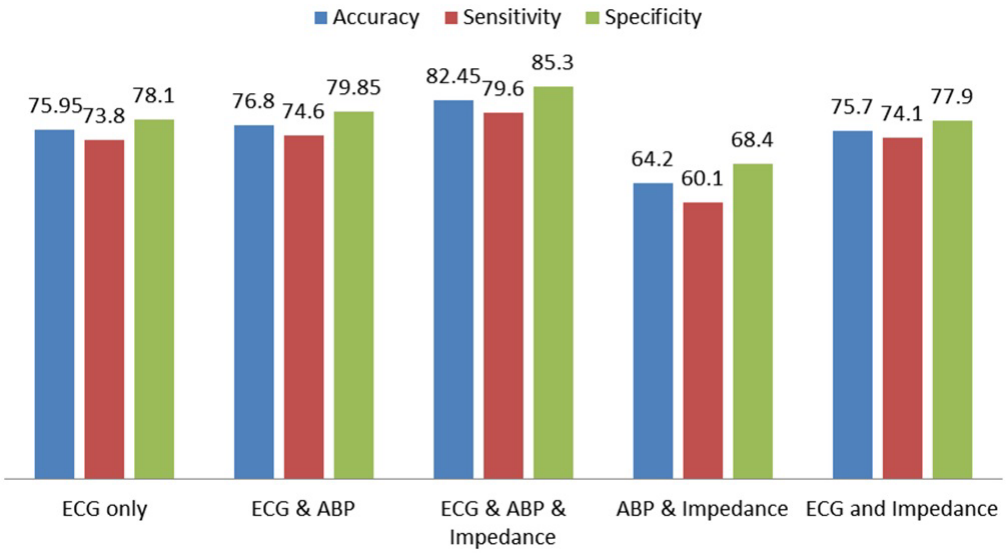


Figure 5.5: Results of the created models using leave one subject out cross validation

5.4 Comparison

The comparison of the algorithm presented in this chapter with other methods is presented in Table 5.3. The same data is used in all studies. The Hakimzadeh et. al. [47] algorithm incorporates the transcranial doppler (TCD) signal only. Ji et. al. method [57] uses ECG, ABP, impedance from the throat (IZT) and from the chest (DZT). In both methods, the best results is reported using SVM classifier. Comparison results are illustrated in Table 5.3.

Table 5.3: Comparison results for the implemented blood loss detection method and other studies. The number of classes is 3. The same mapping and 10-fold cross validation are adopted to report the results in all methods

Method	Se (%)	Sp (%)	Accuracy (%)
This work	79.6	85.3	91.6
Hakimzadeh et. al. [47]	N/A	N/A	70%
Ji et al. [57]	79.8	83	82

5.5 Summary

This chapter presents a method for prediction of the severity of loss of blood volume by incorporating multiple physiological signals such as ECG, ABP and impedance. The method combines features from time domain, frequency domain and arrhythmia. Five models are provided to support the importance of using multiple signals. The dataset used to investigate the accuracy of the model is LBNP. The MIT-BIH database is used to create a classification model to distinguish three types of heartbeats (normal, premature atrial contraction, and ventricular ectopy beat). The method determines the severity of blood loss as one of three functional classes (mild, moderate and severe) as supported by physician.

CHAPTER 6 System Evaluation on Bodymedia Dataset

6.1 Introduction

In this chapter, an evaluation of the algorithms to predict the severity of blood loss is illustrated on a different dataset; the Bodymedia LBNP dataset. The Bodymedia LBNP dataset compromises of 45 subjects obtained from the same LBNP protocol discussed before. All signals in this dataset are sampled using SenseWear® Pro 3 armband (see Figure 6.1), which is a product designed and built by BodyMedia, Inc.

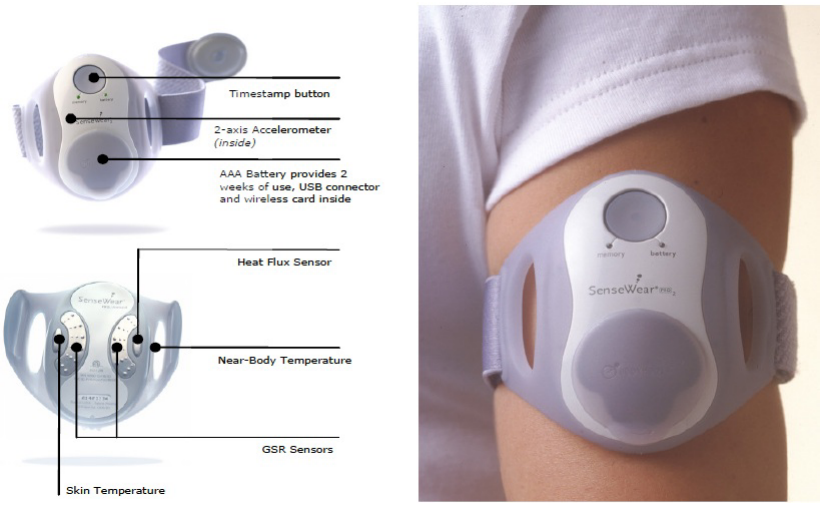


Figure 6.1: SenseWear® pro 3 armband

The armband was designed to be worn under a patient's clothing for continuous data collection and has the following sensors:

1. Two-axis accelerometer to track the movement of the upper arm and body and provides information about body position.

2. A proprietary heat-flux (HF) sensor that measures the amount of heat being dissipated by the body by measuring the heat loss along a thermally conductive path between the skin and a vent on the side of the armband.

3. Skin temperature sensor to measure temperature of the body using sensitive thermistors.

4. Galvanic Skin Response (GSR) sensor that measures the changes in the ability of the skin to conduct electricity caused by sweating or emotional stimuli.

5. ECG sensor to sample the ECG signal with a sampling rate of $128Hz$.

HF, skin temperature and GSR are samples at $1/60Hz$ frequency, whereas the ECG is sampled at $128Hz$.

Since the LBNP signal is proved to simulate blood loss, the armband signals are used to evaluate the algorithms discussed in this dissertation to predict the severity of blood loss. The schematic diagram for the evaluation process is shown in Figure 6.2.

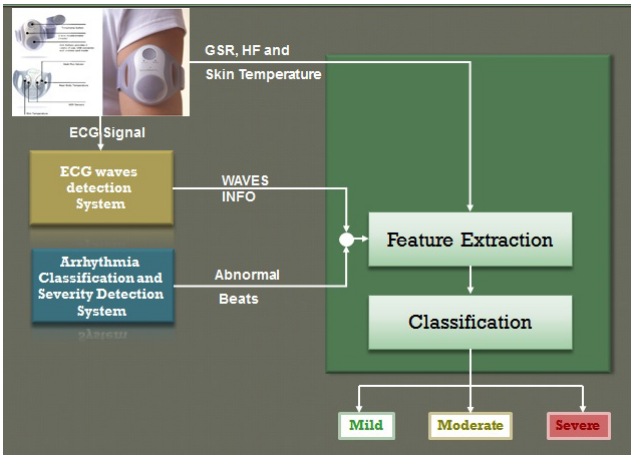


Figure 6.2: Schematic diagram for prediction of loss of blood volume severity using SenseWear® pro 3 armband

The method includes two main steps: Feature Extraction and Classification. The signals that are chosen for analysis are ECG, GSR, HF and skin temperature. Since the former three signals (GSR, HF and skin temperature) are sampled at $1/60Hz$, the system extracts features from them directly without the need to pre-process those signals.

6.2 Methodology

As shown in Figure 6.2, the inputs to this system are from GSR, HF, skin temperature, ECG detection algorithm and arrhythmia classification and severity detection system. Because GSR, HF and skin temperature signals have low sampling frequency, the features are extracted form them directly without pre-processing. The ECG signal is forwarded to the ECG detection algorithm to delineate the five important signal's characteristic points (P, QRS and T). The output of the algorithm is then used to extract features from the time and wavelet domains of the ECG signal and also is forwarded to the arrhythmia system to determine the type of each beat in the signal to extract the rate of abnormal heartbeats. The extracted features are fed into SVM machine learning algorithm to predict the severity of blood loss.

6.2.1 Description of the dataset

As mentioned before, there are 41 subjects in this dataset. Each subject is admitted to the LBNP protocol wearing the armband. The ECG, GSR, HF and skin temperature signals are collected as the pressure is decreased until that subject is unable to continue the protocol and a collapse stage is identified. For compatibility, the same mapping that is presented in Table 5.1 is used in the evaluation process.

6.2.2 Feature Extraction

As shown in Figure 6.2, the inputs for this stage are GSR, HF, skin temperature signals and the ECG signal after it was processed by ECG wave detection system.

The time that a subject expends in the LBNP protocol is divided using a non-overlapping window of size 20 seconds. For example, if a subject stayed for 30 minutes inside the LBNP machine, then the subject's signal is divided into 90 examples.

As discussed before, 20 seconds window is used because the study is aimed to discover a severe loss of blood as soon as possible. Moreover, the window size is believed to be clinically relevant.

From each example, two different sets of features are extracted; time and wavelet domains features. The following subsections describe the extracted features.

6.2.2.1 Time Domain Features

1. ECG Intervals and Amplitude Ratios.

   The ECG signal obtained from the armband is processed by the ECG detection system described in Chapter 3 to detect the characteristic points of each beat in the example. From the detected points of each heartbeat, several clinically important intervals are extracted. These intervals are PQ, QRS, PR, ST and QT. Since there are multiple beats in each window, the standard deviation, mean, median, minimum and maximum of each interval type are calculated and included in the feature set. Moreover, the average ratio of amplitudes are also calculated. These ratios are P-to-Q, Q-to-S, P-to-R, S-to-T and Q-to-T.

   The total number of features generated from the ECG intervals and amplitudes are 30 features.

2. Statistical Measurements.

   From the whole ECG signal, mean, median, standard deviation, minimum and maximum are calculated. In addition, Heart rate variability (HRV) signal is computed from the ECG and the aforementioned measurements are obtained from the generated signal.

   This means that the total of 10 statistical features are extracted from the ECG and HRV signals.

3. Signal Complexity and Mobility.

   Equations number (5.1) and (5.2) are used to find the complexity and mobility of the ECG signal for each example and are included in the feature set. As discussed in Subsection 5.2.3.1, these two features will can quantify the level of variations among the ECG signal. Calculating such features is very important since the heart will pump faster when the body is losing blood in a process called compensation, which means variations in the ECG signal will become more likely.

   There will be 2 features that represent the variation of the ECG signal added to the feature set.

4. KullbackLeibler Distance.

   As mentioned in Subsection 5.2.3.1 , to extract features using Kullback-Leibler Distance (KLD), one need two probability distributions P and Q. Typically P represents the ”true” distribution and Q represents the approximation of P. Equation (5.3) can then be used to compute the KLD between P and Q.

Three subjects are randomly selected and for their ECG signal, the probability distribution function is calculated. The result is three functions; P^1_{ECG}, P^2_{ECG} and P^3_{ECG}. Now for each example, the probability distribution function (Q_{ECG}) for the ECG signal is calculated and equation (5.3) is used to compute $KLD(P^1_{ECG} \,\|\, Q_{ECG})$, $KLD(P^2_{ECG} \,\|\, Q_{ECG})$ and $KLD(P^1_{ECG} \,\|\, Q_{ECG})$.

The above discussion means, that there will be three features that represent the distances between the true distributions (Ps) and the approximate distribution (Q) and are included in the feature set.

5. Rate of Abnormal Heartbeats

   As can be seen in Figure 6.2, the ECG signal is forwarded to the detection system, then the result is forwarded to the feature extraction step and arrhythmia classification and severity detection system. The same arrhythmia system that was discussed in Section 4.3 is used to classify every heartbeat in the ECG signal of each example. The result of this classification will be a vector of classified beats. From this vector, the number of VEBs and PAC are counted and included in the feature set.

   The rate of abnormal beats will add another two features to the feature set.

6. GSR, HF and Skin Temperature Features The sampling rate of the armband to gather GSR, HF and skin temperature is $1/60Hz$, and since each minute of sampling is divided into three windows (each of length 20 seconds). Therefore, GSR, HF and skin temperature signals measured at the beginning of each minute are copied into the second and third windows for that minute.

Example: *Assume that there is a subject, wears an armband, stayed for 2 minutes in the LBNP chamber. The collected signals are then divided into six windows each of length 20 seconds. Because GSR, HF and skin temperature are sampled at* $1/60Hz$*, the first and fourth windows will have a value for these signals. Since these samples are used as features, then the values of GSR, HF and skin temperature signals in the first window are copied into the second and third windows, and the values for the same signals in the fourth window are copied into the fifth and sixth windows.*

Because each example now has only one value for those three signals, no processing is applied to the obtained signals values and are included directly. This will add three features to the feature set to represent GSR, HF and skin temperature.

The total number of features that are extracted from time-domain is 50 features.

6.2.2.2 Wavelet Domain Features

To extract features from the frequency domain, two wavelet transforms are applied to the ECG signal. The first tranformusing daubechie 4 (db4) and the second transform is by dual-tree complex wavelet transform (DT-CWT). The transforms are done at level 4 and the resulted detail coefficients of each level are used to calculate statistical features from the frequency domain.

For each level, the standard deviation, mean, median, kurtosis, and skewness are computed from the detail coefficients of each transform.

The total number of features extracted from wavelet domain is 40 attributes from each example.

6.2.3 Classification

The same machine learning algorithm (SVM) that was used in chapter 5 to predict the severity of blood loss is used. The kernel function in equation (4.2) is adopted for the model. The parameters

are selected based on the optimization tool from LibSvm [12]. The values for the parameters for each model are given in the following section. The total number of features is 90 attributes from time and wavelet domains.

6.3 Results

In this study, 41 subjects from the Bodeymedia LBNP dataset are used. For each stage, a non-normalized and non-overlapping window of size 20 seconds (10000 samples) is defined. Four types of signals (ECG, GSR, HF and skin temperature) are used to extract a total of 90 features. All these signals are non-invasive. The labels for the examples are mild, moderate or severe as defined in Table 5.1. The sensitivity and specificity are calculated using equations (3.6) and (4.4), respectively. The accuracy is calculated using equation (5.5). Two types of validations are considered; 10-folds cross validation and leave one subject out cross validation. The results are illustrated in Tables 6.1 and 6.2.

The true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are calculated and provided in the tables. In addition, the sensitivity (Se), Specificity (Sp) and Accuracy (Acc) are computed and shown in the tables.

Table 6.1: Results of blood loss prediction with 10-folds cross validation on Bodymedia LBNP dataset. The number of cases are 1074 mild, 840 moderate and 758 severe. σ=2^{-1} and $C = 2^5$.

Class	TP	TN	FP	FN	Se	Sp	Acc
Mild	1048	1404	26	103	91.05	98.18	97.58
Moderate	716	1736	124	73	90.75	93.33	85.24
Severe	688	1764	70	44	93.99	96.18	90.77

Figure 6.3 visually illustrates the sensitivity, specificity and accuracy of the model as validated with 10 folds. The values are from Table 6.1.

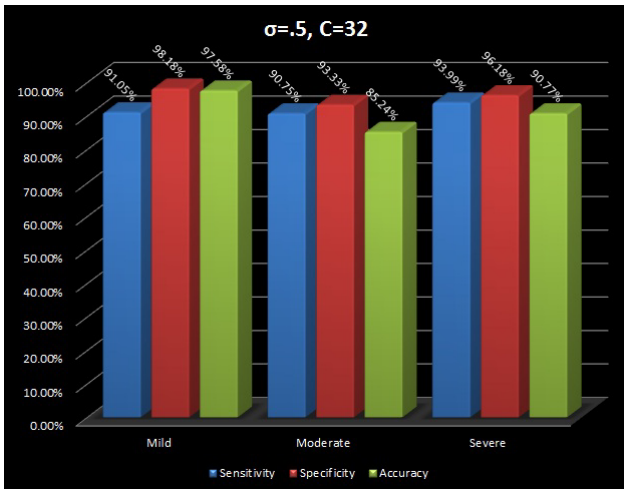


Figure 6.3: Sensitivity, specificity and accuracy of the model with 10 folds cross validation

Table 6.2 shows the results of the model using by subject cross validation as described in Section 5.3. This type of validation is adopted to verify the real accuracy of the model when it is implemented in real systems.

Table 6.2: Results of blood loss prediction with leave one subject out cross validation on Bodymedia LBNP dataset. The number of cases are 1074 mild, 840 moderate and 758 severe. σ=2^{-1} and $C = 2^5$

Class	TP	TN	FP	FN	Se	Sp	Acc
Mild	1038	1181	36	244	80.97	97.04	96.65
Moderate	527	1692	313	104	83.52	84.39	62.74
Severe	654	1565	104	105	86.17	93.77	86.28

Figure 6.4 visually illustrates the sensitivity, specificity and accuracy of the model as validated with leave one subject out is used. The values are in Table 6.2. It is clear that the accuracy dropped to 62.74% when leave one subject out cross validation. This could indicate that the examples generated from one subject are dependable. The only solution to this problem is by increasing the samples in the dataset to adopt different cases.

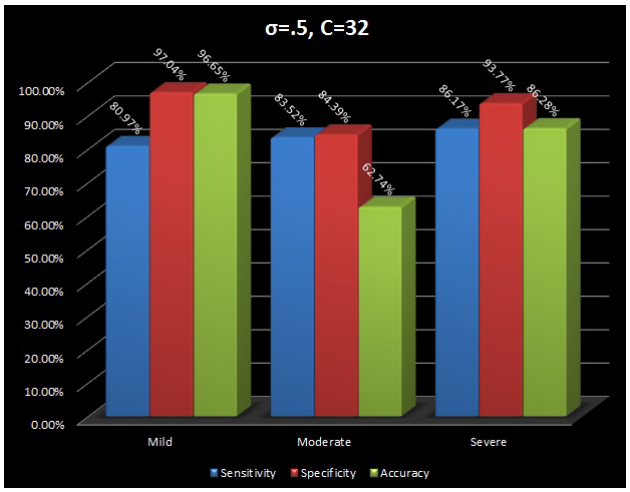


Figure 6.4: Sensitivity, specificity and accuracy of the model with leave one subject out cross validation

Figure 6.5 visually illustrates a comparison between the accuracies obtained from the models that were validated with 10 folds and leave one subject out cross. The results indicate that the highest difference between the two models are in predicting the moderate examples.

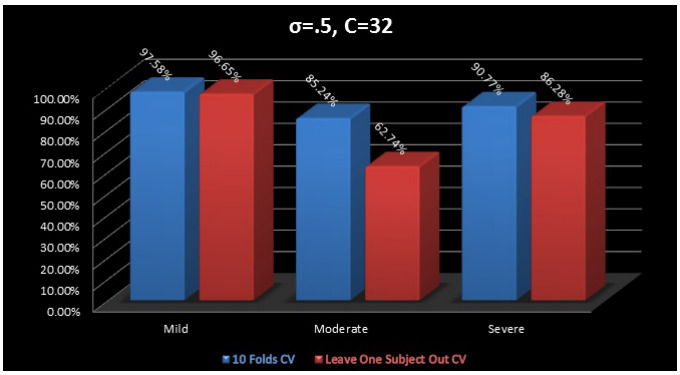


Figure 6.5: Comparison between the accuracy obtained from 10 folds cross validation model and leave one subject out cross validation model

6.4 Summary

This chapter presents an evaluations of the algorithms described in this dissertation on another dataset obtained from Bodymedia, Inc. The data is collected using an armband worn under the clothes as the subject enters a LBNP protocol. Four biomedical signals collected by the armband are used; ECG, GSR, HF and skin temperature. Ninety features are extracted from the time and frequency domain of the signals.

CHAPTER 7 Analysis of Time Complexity

In this chapter, a simplified analysis to approximate the Time-complexity for each system implemented in this dissertation are elaborated. There are three systems created in order to assess the severity of blood loss; ECG detection, Arrhythmia classification and severity detection and blood loss severity prediction. All the complexities are based on Big-O notation, which defines the worst case scenario.

This chapter is organized in four sections. Section 7.1 defines the notation used for the analysis. Section 7.2 elaborates on the time complexity of the ECG detection system. The arrhythmia classification and severity detection system are analyzed in Section 7.3. The computation complexity for blood loss system is outlined in Section 7.4. The chapter is summarized in Section 7.5.

7.1 The "Big-Oh" Notation

In this dissertation, an asymptotic notation called Big-O notation is used to define an upper bound on the worst case scenario for a given algorithm.

Definition Let $f(n)$ be a function that approximate the worst case running time of an algorithm of input size n. Let $g(n)$ be a function mapping nonnegative integers to real numbers. We say that $f(n)$ is $O(g(n)$, if there exist a constant $C > 0$ and an integer constant $n_0 > 0$ such that $f(n) \leq cg(n)$ for sufficiently large $n \geq n_0$.

This means that f is asymptotically upper bounded by g. The definition is often referred to as the "big-oh" notation. Alternatively, we can also say "$f(n)$ is order $g(n)$". In the following sections, an approximation of $f(n)$ is analyzed and the corresponding $g(n)$ is proved using Theorem 7.1.1.

Theorem 7.1.1 (Asymptotic Upper Bound) *Let f and g be two functions that*

$$\lim_{n\to\infty} \frac{f(n)}{g(n)}$$

exists and is equal to some number $c > 0$. Then $f(n) = O(g(n))$.

7.2 Time Complexity Analysis of the ECG Detection System

As can be seen in Figure 3.2, the ECG detection system is compromised of four main parts: Preprocessing, QRS detection, P detection and T detection. Each part contains two components. The following sections will elaborates the complexities of the components that make up the system.

7.2.1 Preprocessing

In this step, two components collaborate to generate artifacts-free signal. The first phase is to apply a band pass Butterworth filter and the second step is to subtract the best-fit line that fits the signal's samples in a window of size equal to the sampling frequency of the signal.

The growth rate function in terms of time when applying a band-pass Butterworth filter is $f(n) = n$, where n is the length of the signal. The complexity of such a filter does not depend on the order of that filter. Even though the order of the filter is n. This means that the growth rate is linear as the order is increased.

The second step is to find the best-fit line and subtract it from the samples of the signal in a window of size equals to the sampling frequency s. Equation 3.1 shows that the growth of the line calculation is linear, hence the complexity to construct such a line is $f(k) = k$, where k is the length of vector x in equation (3.1). As said before, a window of size equals to the sampling

frequency is defined for the baseline removal step. Therefore, the growth rate function to remove the baseline drift in one window is $f(s) = s$. This means that if the signal has n samples, then there are $\frac{n}{s}$ windows that will be defined to remove the baseline drift. Therefore, the overall growth rate of this step is $f(n) = \left(s\frac{n}{s}\right)$.

Therefore, the overall rate of growth for this component is $f(n) = 2n$

7.2.2 QRS Detection

In this part of the system, the signal is transformed with dual-tree complex wavelet transform. For all types of wavelet transform, the computational complexity depends mainly on the implementation method. Nowadays, there exist methods that can generate the wavelet transform of a signal in a growth rate of $f(n) = n$, where n is the length of the signal. The transform is done in this dissertation using Matlab 7.7. The growth rate of WT as implemented in this package is $f(n) = n$.

After the transformation is applied the algorithm squares the resulted detail coefficients and a threshold is applied afterwards. As mentioned before, the transformation takes place at level 4. As a characteristic of wavelet transform, the signal is down sampled by 2 for each decomposition level, which means the length of the detail coefficients that will be squared in the implemented algorithm is $\frac{n}{16}$. The square and threshold are linear operations, hence the growth rate of these steps is $f(n) = \left(2 * \frac{n}{16}\right)$.

Therefore, the overall rate of growth for detecting QRS-complex using the implemented method is $f(n) = n + 2 * \frac{n}{16}$.

7.2.3 P and T Detection

To detect P and T waves, two windows are defined of length $0.16s$, where s is the sampling rate. WT is applied at level 2 for this window and the resulted coefficients are also squared and

thresholded. By using the fact that there are $\frac{n}{s}$, this means that the growth rate to detect P and T waves is $f(n) = 2(0.16n + 2 * \frac{0.16n}{4})$.

7.2.4 Overall Complexity

After approximating the computation complexity for the individual components of the system, the overall complexity is calculated by summing up the overhead for the individual parts, since these components are performed sequentially. Therefore, the complexity of the algorithm is:

$$\begin{aligned} f(n) = (2n) + \left(n + 2 * \frac{n}{16}\right) + \left(2\left(0.16n + 2 * \frac{0.16n}{4}\right)\right) \\ = 3n + \frac{n}{8} + 0.32n + 0.16n \\ = 25\frac{n}{8} + 0.48n \end{aligned} \tag{7.1}$$

Assume that $g(n) = n$, according to theorem 7.1.1, the time complexity of $f(n)$ in Equation 7.1 is $O(n)$.

Proof To proof that $f(n)$ in Equation 7.1 is $O(g(n))$, we will apply the limit in Theorem 7.1.1 to find a constant $c > 0$. We have $f(n) = 25\frac{n}{8} + 0.48n$ and $g(n) = n$. That is

$$\lim_{n\to\infty} \frac{f(n)}{g(n)}$$
$$= \lim_{n\to\infty} \frac{3.125n + 0.48n}{n}$$
$$= \lim_{n\to\infty} \frac{3.605n}{n}$$
$$= \lim_{n\to\infty} 3.605$$
$$= 3.605$$

As the proof shows, there is a constant $c > 0$ that satisfy the limit in Theorem 7.1.1. Since n_0 must be integer, we can say the $f(n)$ in Equation 7.1 is $O(n)$, for $n \geq 4$.

In fact, the following theorem holds for all polynomial functions of degree d.

Theorem 7.2.1 (Asymptotic Upper Bound) *Let f be a polynomial of degree d, such that*

$$f(n) = a_d n^d + a_{d-1} n^{d-1} + ... + a_0$$

where a_d is positive. Then $f = O(n^d)$.

This theorem will be used to find the time complexity in the following sections.

7.3 Time Complexity Analysis of the Arrhythmia Classification and Severity Detection System

7.3.1 Analysis of Arrhythmia Classification System

The arrhythmia classification system is created by training a SVM model using features extracted from time and frequency domains. The computational complexity to train a model using SVM takes very long time. However, once the model is trained, classifying a new beat is easily done by applying Equation 4.3. In fact, the model should be trained off-line, and then it can be saved to classify the type of a new heartbeat. The only concern is how much time is needed to extract

the features from each beat. As can be seen in Figure 4.1, time domain features are obtained from the results of the ECG detection system. Let m denotes the length of the beat that the system will extract the features from. Therefore, the growth rate to extract features from the time domain is $f(m) = 3.605m$ as shown in Equation 6.1. On the other hand, frequency domain features are extracted by applying wavelet transform to the beat at level 4. As a result, the rate of growth to generate the frequency domain features is $f(m) = m$.

Therefore, the overall growth rate to extract the features from a beat is $f(m) = 4.605m$.

7.3.2 Analysis of Arrhythmia Severity Detection System

After classifying ever beat in the ECG, the detection of arrhythmia severity is performed by manipulating a vector from the classified beats. Let c denotes the number of beats in the vector. A window of the 30 last beats is defined and the deterministic finite automate (DFA) is used to test this window. If the DFA did not result in a severe arrhythmia according to the rules, the window is shifted five beats and the DFA is tested again. That means, there are $(c - 30)$ windows that will be tested. Since the analysis of the vector is sequential, then the growth rate to process one window is 30. Therefore, the overall time growth rate to test the whole vector is $f(c) = 30(c-30) = 30c-900$.

7.3.3 Overall Complexity

According to the decision before, and since the classification takes place first then the detection, the overall time complexity of the arrhythmia classification and severity detection system using Theorem 7.2.1 is:

$$= f(m) + f(c)$$

$$= 2m + 30c - 900$$

$$= O(c)$$

For sufficiently large $c \gg m$.

7.4 Time Complexity Analysis of the Blood Loss Prediction System

As Figure 5.1 shows, the blood loss prediction system uses information from the ECG detection system and the arrhythmia classification system. The system creates a SVM model, and as said before, the model should be trained off-line, stored, and then used to test other signals. There are five models implemented in this dissertation to highlight the importance of incorporating different types of signals to assess the severity of blood loss. From each considered signal, different types of features are extracted. Though, the analysis of time required to extract the features is provided from the signal type point of view. As a reminder, a window of size 20 seconds is used as one example in this system. Let n defines the length of the window, and b represents the number of beats in that window.

7.4.1 Time Complexity Analysis when the ECG Signal is Used

Figure 5.1 shows that the ECG signal is passed through the ECG wave detection system and the arrhythmia classification system and the results are forwarded to a feature extraction step. Two types of features are calculated as follows:

7.4.1.1 Time Domain Features

1. ECG Intervals and Amplitude Ratios.

As discussed in Section 5.2.3.1, the intervals and amplitudes features are extracted from the output of the ECG detection system. Equation (7.1) shows that $f(n)$ of the detection system is $f(n) = 3.605n$, where n is the length of the window. A total of 20 features are extracted by averaging the information from the ECG deflection points. Since there are b heartbeats in the window, then the growth rate for generating such features is $f(n, b) = 3.605n + 20b$. That is, we need $3.605n$ to detect the waves, and $20b$ to calculate the features from the results.

2. Statistical Measurements

As well as the ECG signal, HRV signal is computed from the results of the ECG detection system and the mean, standard deviation, and median are calculated from the two signals (ECG and HRV). Since b represents the number of beats in the window, then it takes $f(b) = b - 1$ to generate the HRV signal. The statistical measurements considered in this study are lucid and linear. A total of 6 features are calculated from ECG and HRV. Therefore, the overall rate of growth to find the statistical attributes is $f(n, b) = 3n + 3(b - 1) = 3n + 3b - 3$.

3. Signal Complexity and Mobility

These are two features that are used to quantify the first and second order variations of the signal. Equations (5.1) and (5.2) are used to calculated the complexity and mobility of the ECG signal. It takes $n - 1$ to find vector d and $n - 2$ to find vector g. For S_0, it is clear that we need n steps, and to calculate S_1 and S_2, they take $n - 1$ and $n - 2$. Therefore, the growth rate extract the signal complexity from Equation (5.1) is $f(n) = n - 2 + n - 1 + n = 3n - 3$, and $f(n)$ to compute the mobility of a signal is $f(n) = n - 1 + n = 2n - 1$. Then, the overall growth rate to find the mobility and complexity of a signal is $f(n) = 3n - 3 + 2n - 1 = 5n - 4$.

4. Kullback Leibler Distance

   To calculate features from Kullback Leibler distance (KLD), three subjects are picked randomly and the histogram for their ECG is calculated. These histogram represent the "true" distributions P^i, for $i = 1, 2, 3$. In fact, these histograms are calculated off-line, hence the time incurred is insignificant. The histogram for the window is calculated to serve as the approximation distribution Q of P. Calculating the histogram of a signal is linear and $f(n)$ is equal to n, since we simply count the frequency of each amplitude in the digitized signal. Equation (5.3) is then used to calculated KLD between P^i and Q. Clearly, the equation is linear, therefore the growth rate to compute the KLD features from the ECG is $f(n) = n + 3n = 4n$. That is, it takes n to construct the histogram, and another n to find $KLD(P^i||Q)$, for $i = 1, 2, 3$.

5. Rate of Abnormal Heartbeats

   These two features are calculated from the arrhythmia classification system. It is shown before that the time complexity for this system is $O(2m)$, where m is the length in time of a heartbeat. To extract these features, the system should classify every beat in the defined window. Let k represents the number of heartbeats that the window has, therefore, the rate of growth to calculate the rate of abnormal heartbeats is $f(k, m) = (k(2m))$. It is clear that the multiplication of the number of heartbeats by the length of longest heartbeat is greater that or equal to the length of the signal. That is $km \geq n$. Though, the rate of growth to find the abnormal heartbeats is $f(n) = 2n$.

7.4.1.2 Wavelet Domain Features

Wavelet features are extracted by applying wavelet transformation at level 4 as discussed in Section 5.2.3.2. Standard deviation, mean, median, Kurtosis, and Skewness are extracted from the detail coefficients for each level of the transformation. As mentioned before, the transformation costs $O(n)$ and the resulted coefficients are down sampled by 2 for each decomposition level. Therefore, these simple features take $f(n) = O(n) + 5n/2 + 5n/4 + 5n/8 + 5n/16$. That is, it takes $O(n)$ to generate the transformation and the time to calculate the five features from each level is $5n/2^i$ for $i = 1, 2, 3, 4$.

In addition, the entropy of the transformation is also calculated using equation (5.4). To calculate p for the equation, the approximation and detail coefficients are used. Since the number of the approximation coefficients at level 4 is $n/16$. Therefore, to calculate the entropy of the transformation it takes $f(n) = n/16 + n/2 + n/4 + n/8 + n/16$. That is, to find the summation of the approximation coefficients it takes $n/16$, and the summation of the detail coefficients at level i takes $n/2^i$, for $i = 1, 2, 3, 4$. The overall growth rate to extract features from wavelet domain is

$$\begin{aligned} f(n) &= O(n) + 5n/2 + 5n/4 + 5n/8 + 5n/16 + n/16 + n/2 + n/4 + n/8 + n/16 \\ &= O(n) + 6n/2 + 6n/4 + 6n/8 + 7n/16 \end{aligned}$$

7.4.1.3 Overall Complexity when ECG is used

To find the growth rate to extract the whole features from the window, we can simply sum up the individual growth rates for each type of feature. That is, the overall growth rate is

$$f(n,b) = 3.605n + 20b + 3n + 3b - 3 + 5n - 4 + 4n + 2n + O(n) + 6n/2 + 6n/4 + 6n/8 + 7n/16$$
$$= 23.2925n + 23b - 7$$

Finally, referring to Theorem 7.2.1, the time complexity to generate features from the ECG signal to create a blood loss prediction system is $O(n)$.

7.4.2 Time Complexity Analysis when the ABP or Impedance Signals are Used

Since the same set of features are extracted from the ABP or impedance signals, this section will analyze the time complexity when one of them is incorporated. As Figure 5.1 shows, each signal is preprocessed by applying a suitable filter to reduce the effect of noise and other artifacts. The applied filter is a Butterworth band pass filter as discussed in Section 5.2.2. This type of filter takes no more than $O(n)$ and the complexity has nothing to do with the order of the filter.

The growth rate function to extract features from each signal is analyzed in the following sections.

7.4.2.1 Time Domain Features

1. Statistical Measurements The mean, standard deviation, and median are calculated from both signals. As said before, it takes n steps to compute each feature. A total of 3 features are calculated from each signal. Therefore, the overall rate of growth to find the statistical attributes of a signal is $f(n) = 3n$.

2. Signal Complexity and Mobility

Section 6.3.1.1. shows that the overall growth rate to find the mobility and complexity of a signal is $f(n) = 3n - 3 + 2n - 1 = 5n - 4$.

3. Kullback Leibler Distance

   The growth rate to compute the KLD features from ABP or Impedance is $f(n) = n+3n = 4n$. That is, it takes n to construct the histogram, and another n to find $KLD(P^i||Q)$, for $i = 1, 2, 3$. This is discussed in Section 6.3.1.1.

7.4.2.2 Wavelet Domain Features

The same wavelet features that are extracted from the ECG signal are also extracted from ABP or Impedance. That means, the growth rate is the same. Section 6.3.1.2 shows that the overall growth rate to extract features from wavelet domain is $f(n) = O(n) + 6n/2 + 6n/4 + 6n/8 + 7n/16$.

7.4.2.3 Overall Complexity when ABP or Impedance are used

To find the growth rate to extract the whole features from the window, we can simply sum up the individual growth rates for each type of feature. So, the overall growth rate is.

$$\begin{aligned} f(n) &= 3n + 5n - 4 + 4n + 6n/2 + 6n/4 + 6n/8 + 7n/16 + O(n) \\ &= 17.6875n - 4 + O(n) \end{aligned}$$

Referring to Theorem 7.2.1, the time complexity to incorporate features from ABP or Impedance signals in a blood loss prediction system is $O(n)$.

7.5 Summary

In this chapter, analysis of time complexity of each system implemented in this dissertation is discussed. The provided analysis is an approximation and the running time may increase by a constant factor. The discussion shows that the algorithms implemented are simple and take $O(n)$. Therefore, the methods can be implemented in real-time.

CHAPTER 8 Conclusions and Future Work

In this dissertation, a system that predicts the severity of blood volume loss by incorporating multiple physiological signals is presented. The majority of previous studies concentrate on analyzing the complexity of heart rate variability (HRV) by using traditional methods such as power spectrum density (PSD) and fractal dimension. However, these methods are only applicable to signals that have stationary property (the statistical measurements are the same at all times). In this study, wavelet transform (WT) is used to process non-stationary signals such as ECG, ABP, impedance and HRV.

This chapter is organized in two sections. Section 8.1 provides the conclusions of the different systems outlined in this dissertation. Plans for the future work are proposed in Section 8.2.

8.1 Conclusions

The system implemented is compromised of three subsystems. The conclusions for each subsystem is given below:

8.1.1 Conclusions on the ECG Detection System

Prior to prediction, ECG characteristic points (P-QRS-T) are detected using a novel approach based on the physiology of the ECG. This system is implemented to extract informative features from not only QRS complex, but also P and T waves.

The results of the system show that it is reliable and robust to noise or artifacts exist in the ECG signal. It is important to note that some of the previous studies exclude records number 107,

201 and 231, since they have heavy noise in the signal obtained. However, when reporting the results of the system, these subjects are included.

In addition to the detection of QRS as well as P and T waves, the algorithm described can distinguish weather QRS has positive or negative deflection. This is important to consider when the application that incorporates such a system needs this information, specifically, for arrhythmia classification.

These properties of the detection system make it novel and applicable to a wide spectrum of applications where the ECG signal is incorporated.

8.1.2 Conclusions on the Arrhythmia Classification and Severity Detection System

The novelty of the created system is to detect the severity of arrhythmia encapsulated in the ECG signal's heartbeats using several simple rules supported by the medical literature. The rules are implemented with a deterministic finite automata (DFA) that runs the expression from the heartbeats after classification. Such a system, provided by the results, is capable to distinguish between severe and benign arrhythmias.

The implemented algorithm is scalable. This is proved when the trained model that uses annotated beats from the MIT/BIH arrhythmia database is incorporated to classify heartbeats of the LBNP dataset to extract the rate of abnormal beats as a feature for the blood loss severity prediction system. The size of the trained model is small, when saved as a JAVA serializable object (¡ 700 KBs). The running time of the algorithm is polynomial, since the features are extracted from time and frequency domains by transforming the beat using a wavelet transformation method, which takes $O(n)$ in the worst case.

The method described will allow for the development of a computationally inexpensive decision support system to rapidly detect and report ectopy in an early warning manner that may allow

health care providers more time to react prior to possible patient deterioration.

8.1.3 Conclusions on the Prediction of Blood Volume Loss System

The system focuses specifically on manipulating multiple physiological signals that are easy and simple to acquire using remote devices to extract informative features to detect the severity of hemorrhage. The dissertation presents a novel work in integrating several biomedical signals such as ECG, ABP and impedance, to automatically classify the severity of blood volume loss into one of three functional classes (mild, moderate and severe).

The results provided show that, when incorporating three signals (ECG, ABP and Impedance) the accuracy is very close to the ECG and Impedance signal pairs. However, this must be elaborated from two different points of view. From medical perspective, incorporating three signals is more useful than processing two of them. The results illustrate that there are 12 objects will be classified correctly, when three different types of signals are used. Keep in mind that these are 12 windows, where each window is 20 seconds. This means that there will be an addition of about four minutes for a physician to make a decision ($4 = 12 * 20/60$). This is very important time frame in high pace environments, where physicians should take critical decisions in very short amount of time. However, from technical point of view, incorporating only two signals is more feasible from time and space complexities.

8.2 Future Work

The first immediate goal is to obtain expert assistance from multiple cardiologist in labeling the datasets. Although, such a need is likely to prove difficult logistically due to the expected volume of data and the time demands involved. More rigorous testing demands a gold standard, which in this dissertation is manual for detecting the ECG deflection points and the accuracy of applying the

arrhythmia classification model over different dataset cannot be obtained because the heartbeats of the LBNP dataset are not labeled.

The most crucial stage of the future work is in integrating more information from the arrhythmia system. In this dissertation, informative features such as the number of ectopic beats per window are calculated and combined to the set of features extracted from the ECG, ABP and Impedance signals. This feature, i.e. number/rate of ectopic beats, is believed to be correlated to hypoxia. The relationship between hypoxia and arrhythmia is explained on the metabolic basis [19, 71, 101, 129]. It is anticipated that adding the rate of other ectopic beats that are not considered in this study is crucial and may increase the accuracy of the hemorrhage predicition system.

Therefore, the second aim is to expand the work done for arrhythmia classification and severity detection system to obtain different abnormal beat types and include the rate for each of them in the system implemented in this dissertation. I believe that the blood loss severity prediction system can be enhanced substantially, when the arrhythmia classification and severity detection system is used to extract more than two types of abnormal heartbeats. However, considering extracting more than two types will directly affect the complexity of an arrhythmia system. Moreover, the accuracy of the system is affected by the morphology of some of the ectopy that will fool the algorithm.

Concerning the technical aspects, an important future task is to improve the robustness of the overall processes. This will include testing the ECG deflection points detection algorithm with several signals from the wide spectrum of different beat patterns. The obtained results should be evaluated by a cardiology to asses the capability of the implemented algorithm.

Bibliography

[1] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. *Machine Learning: ECML 2004*, pages 39–50, 2004.

[2] RV Andreao, B. Dorizzi, and J. Boudy. ECG signal analysis through hidden Markov models. *IEEE Transactions on Biomedical Engineering*, 53(8):1541–1549, 2006.

[3] M. Arif, MU Akram, and FA Afsar. Arrhythmia beat classification using pruned fuzzy k-nearest neighbor classifier. In *Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of*, pages 37–42. IEEE, 2009.

[4] S. Bahrami, K. Zimmermann, Z. Szelényi, J. Hamar, F. Scheiflinger, H. Redl, and W.G. Junger. Small-volume fluid resuscitation with hypertonic saline prevents inflammation but not mortality in a rat model of hemorrhagic shock. *Shock*, 25(3):283, 2006.

[5] A.I. Batchinsky, W.H. Cooke, T.A. Kuusela, B.S. Jordan, J.J. Wang, and L.C. Cancio. Sympathetic nerve activity and heart rate variability during severe hemorrhagic shock in sheep. *Autonomic Neuroscience*, 136(1-2):43–51, 2007.

[6] J.J. Bazarian, J. Mcclung, M.N. Shah, Y. Ting Cheng, W. Flesher, and J. Kraus. Mild traumatic brain injury in the United States, 1998-2000. *Brain Injury*, 19(2):85–91, 2005.

[7] R.F. Bellamy, P.A. Maningas, and J.S. Vayer. Epidemiology of trauma: military experience. *Annals of Emergency Medicine*, 15(12):1384–1388, 1986.

[8] Y. Ben-Menachem, D.M. Coldwell, J.W. Young, and A.R. Burgess. Hemorrhage associated with pelvic fractures: causes, diagnosis, and emergent management. *American J. Roentgenology*, 157(5):1005–1014, November 1991.

[9] DS Benitez, PA Gaydecki, A. Zaidi, and AP Fitzpatrick. A new QRS detection algorithm based on the Hilbert transform. *Computers in Cardiology 2000*, pages 379–382, 2000.

[10] S. Bilgin, O.H. Çolak, E. Koklukaya, and N. ArI. Efficient solution for frequency band decomposition problem using wavelet packet in HRV. *Digital Signal Processing*, 18(6):892–899, 2008.

[11] RJ Bolton and LC Westphal. ECG display and QRS detection using the Hilbert Transform. *Computers in Cardiology, IEEE Computer Society, Washington, DC*, pages 463–466, 1985.

[12] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[13] L. Chen, T.M. McKenna, A.T. Reisner, A. Gribok, and J. Reifman. Decision tool for the early diagnosis of trauma patient hypovolemia. *Journal of biomedical informatics*, 41(3):469–478, 2008.

[14] WT Cheng and KL Chan. Classification of electrocardiogram using hidden Markov models. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, pages 143–146, 1998.

[15] F. Chiarugi, V. Sakkalis, D. Emmanouilidou, T. Krontiris, M. Varanini, and I. Tollis. Adaptive threshold QRS detector with best channel selection based on a noise rating system. *Computers in Cardiology, 2007*, pages 157–160, 2007.

[16] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury. Hidden Markov tree modeling of complex wavelet transforms. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, volume 1, pages 133–136. IEEE, 2000.

[17] I.I. Christov. Real time electrocardiogram QRS detection using combined adaptive threshold. *BioMedical Engineering OnLine*, 3(1):28, 2004.

[18] K.J. Cios, W. Pedrycz, and R. Świniarski. *Data mining methods for knowledge discovery*. Kluwer Academic Publishers, 1998.

[19] R.E. CLARK, I. CHRISTLIEB, M. SANMARCO, R. DIAZ-PEREZ, J. DAMMANN, et al. Relationship of Hypoxia to Arrhythmia and Cardiac Conduction Hemorrhage: An Experimental Study. *Circulation*, 27(4):742, 1963.

[20] D.A. Coast and G.G. Cano. QRS detection based on hidden Markov modeling. page 34, 1989.

[21] DA Coast, RM Stern, GG Cano, and SA Briller. An approach to cardiac arrhythmia analysis using hidden Markov models. *IEEE Transactions on biomedical Engineering*, 37(9):826–836, 1990.

[22] VA Convertino. Lower body negative pressure as a tool for research in aerospace medicine and military medicine. *J Gravit Physiol*, 8:1–14, 2001.

[23] W.H. Cooke, K.L. Ryan, and V.A. Convertino. Lower body negative pressure as a model to study progression to acute hemorrhagic shock in humans. *Journal of Applied Physiology*, 96(4):1249, 2004.

[24] W.H. Cooke, J. Salinas, V.A. Convertino, D.A. Ludwig, D. Hinds, J.H. Duke, F.A. Moore, and J.B. Holcomb. Heart rate variability and its association with mortality in prehospital trauma patients. *The Journal of Trauma*, 60(2):363, 2006.

[25] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[26] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research*, 2:265–292, 2002.

[27] K. Das, B. Giesbrecht, and M.P. Eckstein. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. *Neuroimage*, 51(4):1425–1437, 2010.

[28] P. De Chazal, M. O'Dwyer, and RB Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7):1196–1206, 2004.

[29] P. De Chazal and RB Reilly. Automatic classification of ECG beats using waveform shape and heart beat interval features. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)*, volume 2, 2003.

[30] D. Demetriades, B. Kimbrell, A. Salim, G. Velmahos, P. Rhee, C. Preston, G. Gruzinski, and L. Chan. Trauma Deaths in a Mature Urban Trauma System: Is Trimodal Distribution a Valid Concept? *Journal of the American College of Surgeons*, 201(3):343–348, 2005.

[31] S. Dikmen, A. McLean, and N. Temkin. Neuropsychological and psychosocial consequences of minor head injury. *British Medical Journal*, 49(11):1227, 1986.

[32] Z. Dokur and T. Olmez. ECG beat classification by a novel hybrid neural network. *Computer methods and programs in biomedicine*, 66(2-3):167–181, 2001.

[33] P.L. Dragotti and M. Vetterli. Wavelet footprints: Theory, algorithms, and applications. *Signal Processing, IEEE Transactions on*, 51(5):1306–1323, 2003.

[34] B. J. Eastridge, A. M. Starr, J. P. O'Keefe, and E. Grant. The importance of fracture pattern in guiding therapeutic decision-making in patients with hemorrhagic shock and pelvic ring disruptions. *J. Trauma*, 53(3):446–450, September 2002.

[35] M. Engin. ECG beat classification using neuro-fuzzy network. *Pattern Recognition Letters*, 25(15):1715–1722, 2004.

[36] J. Englander, K. Hall, T. Stimpson, and S. Chaffing. Mild traumatic brain injury in an insured population: subjective complaints and return to employment. *Brain Injury*, 6(2):161–166, 1992.

[37] T.P. Exarchos, M.G. Tsipouras, C.P. Exarchos, C. Papaloukas, D.I. Fotiadis, and L.K. Michalis. A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. *Artificial Intelligence in medicine*, 40(3):187–200, 2007.

[38] Centers for Disease Control and Prevention. Web-based inquiry statistics query and reporting system (wisqars). national center for injury prevention and control, centers for disease control and prevention, 2007. http://www.cdc.gov/injury/ wisqars/index.html (accessed feb ruary 2010).

[39] Centres for Disease Control and Prevention. Facts about concussion and brain injury. 2004.

[40] GM Friesen, TC Jannett, MA Jadallah, SL Yates, SR Quint, and HT Nagle. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1):85–98, 1990.

[41] D. Ge, N. Srinivasan, and S.M. Krishnan. Cardiac arrhythmia classification using autoregressive modeling. *BioMedical Engineering OnLine*, 1(1):5, 2002.

[42] A. Ghaffari and MR Homaeinezhad. *Peak detection via modified Hilbert transform: application to QRS and end-systolic end-diastolic events detection*. PhD thesis, MSc Thesis, Department of Mechanical Engineering, KN Toosi University of Technology, Tehran, Iran, 2006.

[43] A. Ghaffari and MR Homaeinezhad. *Peak detection via modified Hilbert transform: application to QRS and end-systolic end-diastolic events detection*. PhD thesis, MSc Thesis, Department of Mechanical Engineering, KN Toosi University of Technology, Tehran, Iran, 2006.

[44] S. Graja, J.M. Boucher, E.N.S. des Telecommun, and F. Brest. Hidden Markov tree model applied to ECG delineation. *IEEE Transactions on Instrumentation and Measurement*, 54(6):2163–2168, 2005.

[45] I. Guler et al. ECG beat classifier designed by combined neural network model. *Pattern Recognition*, 38(2):199–208, 2005.

[46] G. Gutierrez, H.D. Reines, and M.E. Wulf-Gutierrez. Clinical review: Hemorrhagic shock. *CRITICAL CARE-LONDON-*, 8:373–381, 2004.

[47] R. Hakimzadeh, S.Y. Ji, R. Smith, K. Ward, K. Daneshvar, and K. Najarian. Processing of

transcranial doppler for assessment of blood volume loss. In *Information Reuse & Integration, 2009. IRI'09. IEEE International Conference on*, pages 6–10. IEEE, 2009.

[48] P.S. Hamilton and W.J. Tompkins. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Transactions on Biomedical Engineering*, pages 1157–1165, 1986.

[49] C.R. Heier, R. Satta, C. Lutz, and C.J. DiDonato. Arrhythmia and cardiac defects are a feature of spinal muscular atrophy model mice. *Human molecular genetics*, 19(20):3906, 2010.

[50] E.K. Heist and J.N. Ruskin. Drug-Induced Arrhythmia. *Circulation*, 122(14):1426, 2010.

[51] Y.H. Hu, S. Palreddy, and W.J. Tompkins. A patient-adaptable ecg beat classifier using a mixture of experts approach. *Biomedical Engineering, IEEE Transactions on*, 44(9):891–900, 1997.

[52] EE Jaeggi, JC Fouron, and SP Drblik. Fetal atrial flutter: diagnosis, clinical features, treatment, and outcome. *The Journal of pediatrics*, 132(2):335–339, 1998.

[53] S. Jankowski, A. Oreziak, A. Skorupski, H. Kowalski, Z. Szymanski, and E. Piatkowska-Janko. Computer-aided morphological analysis of holter ECG recordings based on support vector learning system. *Computers in Cardiology, 2003*, pages 597–600, 2003.

[54] I. Jekova, G. Bortolan, and I. Christov. Pattern recognition and optimal parameter selection in premature ventricular contraction classification. *Computers in Cardiology, 2004*, pages 357–360, 2004.

[55] T.J. Jensen, J. Haarbo, S.M. Pehrson, and B. Thomsen. Impact of premature atrial contractions in atrial fibrillation. *Pacing and clinical electrophysiology*, 27(4):447–452, 2004.

[56] A. Jervell and F. Lange-Nielsen. Congenital deaf-mutism, functional heart disease with prolongation of the QT interval, and sudden death. *American Heart Journal*, 54(1):59–68, 1957.

[57] Soo-Yeon Ji, Wenan Chen, K. Ward, C. Rickards, K. Ryan, V. Convertino, and K. Najarian. Wavelet based analysis of physiological signals for prediction of severity of hemorrhagic shock. *Complex Medical Engineering, 2009. CME. ICME International Conference on*, pages 1 –6, april 2009.

[58] Soo-Yeon Ji, Rebecca Smith, Toan Huynh, and Kayvan Najarian. A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries. *BMC Medical Informatics and Decision Making*, 9(1):2, 2009.

[59] S.Y. Ji, A.A.R. Bsoul, K. Ward, K. Ryan, C. Rickard, V. Convertino, and K. Najarian. Abstract P195: Incorporating Physiological Signals to Blood Loss Prediction Based on Discrete Wavelet Transformation. *Circulation*, 120(18 Supplement):S1483, 2009.

[60] S.Y. Ji, W. Chen, K. Ward, C. Rickards, K. Ryan, V. Convertino, and K. Najarian. Wavelet based analysis of physiological signals for prediction of severity of hemorrhagic shock. In *Complex Medical Engineering, 2009. CME. ICME International Conference on*, pages 1–6. IEEE.

[61] V. Kecman. Basics of Machine Learning by Support Vector Machines. *Real World Applications of Computational Intelligence*, 179:49–103, 2005.

[62] D.M. Kelley. Hypovolemic shock: an overview. *Critical Care Nursing Quarterly*, 28(1):2, 2005.

[63] JGC Kemmelings, AC Linnenbank, SLC Muilwijk, A. SippensGroenewegen, A. Peper, and CA Grimbergen. Automatic QRS onset and offset detection for body surface QRSintegral mapping of ventricular tachycardia. *IEEE Transactions on Biomedical Engineering*, 41(9):830–836, 1994.

[64] S.Y. Kim, M.J. Kim, S.M. Cadarette, and D.H. Solomon. Bisphosphonates and risk of atrial fibrillation: a meta-analysis. *Arthritis Res Ther*, 12:R30, 2010.

[65] T. Klingenheben, C. Sticherling, M. Skupin, and S.H. Hohnloser. Intracardiac QRS electrogram width an arrhythmia detection feature for implantable cardioverter defibrillators: exercise induced variation as a base for device programming. *Pacing and Clinical Electrophysiology*, 21(8):1609–1617, 2006.

[66] C.M. Kloppel, S.and Stonnington, J. Barnes, F. Chen, C. Chu, C.D. Good, I. Mader, L.A. Mitchell, A.C. Patel, and C.C. others Roberts. Accuracy of dementia diagnosis - a direct comparison between radiologists and a computerized method. *Brain Injury*, 131(11):2969, 2008.

[67] M.N. KOTLER, B. TABATZNIK, M.M. MOWER, and S. TOMINAGA. Prognostic significance of ventricular ectopic beats with respect to sudden death in the late postinfarction period. *Circulation*, 47(5):959, 1973.

[68] M. Kubat, R. Holte, and S. Matwin. Learning when negative examples abound. *Machine Learning: ECML-97*, pages 146–153, 1997.

[69] M. Kubat and S. Matwin. Addressing the curse of imbalanced training sets: one-sided selection. In *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, pages 179–186. Citeseer, 1997.

[70] S. Kullback. The kullback-leibler distance. *The American Statistician*, 41(4):340–341, 1987.

[71] VA Kurien and MF Oliver. A metabolic cause for arrhythmias during acute myocardial hypoxia. *The Lancet*, 295(7651):813–815, 1970.

[72] Y. Kutlu and D. Kuntalp. A multi-stage automatic arrhythmia recognition and classification system. *Computers in biology and medicine*, 41(1):37–45, 2011.

[73] J.A. Langlois, W. Rutland-Brown, and M.M. Wald. The epidemiology and impact of traumatic brain injury: a brief overview. *The Journal of head trauma rehabilitation*, 21(5):375, 2006.

[74] J.H. Lee, J.L. Choi, S.W. Chung, and D.W. Kim. A Survival Prediction Model for Rats with Hemorrhagic Shock Using an Artificial Neural Network. *Journal of the Korean Society of Emergency Medicine*, 21(3):321–327, 2010.

[75] C. Li, C. Zheng, and C. Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1):21–28, 1995.

[76] C.H. Lin, C.L. Kuo, J.L. Chen, and W.D. Chang. Fractal features for cardiac arrhythmias recognition using neural network based classifier. In *Networking, Sensing and Control, 2009. ICNSC'09. International Conference on*, pages 930–935. IEEE.

[77] B. Mac Namee, P. Cunningham, S. Byrne, and O.I. Corrigan. The problem of bias in training data in regression problems in medical decision support. *Artificial Intelligence in Medicine*, 24(1):51–70, 2002.

[78] S.N. Macciocchi, D.B. Reid, and J.T. Barth. Disability following head injury. *Current Opinion in Neurology*, 6(5):773, 1993.

[79] SZ Mahmoodabadi, A. Ahmadian, MD Abolhasani, M. Eslami, and JH Bidgoli. ECG feature extraction based on multiresolution wavelet transform. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 3902–3905, 2005.

[80] M. Malik, J.T. Bigger, A.J. Camm, R.E. Kleiger, A. Malliani, A.J. Moss, and P.J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354, 1996.

[81] R. Mark and G. Moody. Mit-bih arrhythmia database 1997.

[82] Marie-Jocelyne Martel. HEMORRHAGIC SHOCK. *JOGC*, page 1, 2002.

[83] J.P. Martínez, R. Almeida, S. Olmos, A.P. Rocha, and P. Laguna. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on biomedical engineering*, 51(4):570–581, 2004.

[84] V.A. Mashin. Nonstationarity and duration of the cardiac interval time series in assessing the functional state of operator personnel. *Biophysics*, 52(2):241–247, 2007.

[85] J.G. McManus, B.J. Eastridge, C.E. Wade, and J.B. Holcomb. Hemorrhage control research on today's battlefield: lessons applied. *The Journal of Trauma*, 62(6):S14, 2007.

[86] VU Medina, R. González-Camarena, and JC Echeverria. Effect of noise sources on the averaged PQRST morphology. *Computers in Cardiology*, 32:743, 2005.

[87] F. Melgani and Y. Bazi. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. *IEEE Transactions on Information Technology in Biomedicine*, 12(5):667–677, 2008.

[88] K. Minami, H. Nakajima, and T. Toyoshima. Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE transactions on Biomedical Engineering*, 46(2):179–185, 1999.

[89] M.L. Minsky. *Computation: finite and infinite machines*. 1967.

[90] A.J. Moss, J.J. DeCamilla, H.P. Davis, and L. Bayer. Clinical significance of ventricular ectopic beats in the early posthospital phase of myocardial infarction*. *The American journal of cardiology*, 39(5):635–640, 1977.

[91] J. Mourao-Miranda, A.L.W. Bokde, C. Born, H. Hampel, and M. Stetter. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage*, 28(4):980–995, 2005.

[92] K. Najarian and R. Splinter. *Biomedical signal and image processing*. CRC Press, 2006.

[93] J.A. Nasiri, M. Naghibzadeh, H.S. Yazdi, and B. Naghibzadeh. Ecg arrhythmia classification with support vector machines and genetic algorithm. In *Computer Modeling and Simulation, 2009. EMS'09. Third UKSim European Symposium on*, pages 187–192. IEEE, 2009.

[94] Eric W. Nawar., Richard W. Niska., and Jianmin Xu. National Hospital Ambulatory Medical Care Survey: 2005 Emergency Department Summary. *Advance Data From Vital and Health Statistics*, 386(320):1–32, 2007.

[95] M.L. O'Donnell, M. Creamer, P. Pattison, and C. Atkin. Psychiatric morbidity following injury. *American Journal of Psychiatry*, 161(3):507, 2004.

[96] T. Ohe, M. Matsuhisa, S. Kamakura, J. Yamada, I. Sato, K. Nakajima, and K. Shimomura. Relation between the widening of the fragmented atrial activity zone and atrial fibrillation. *The American journal of cardiology*, 52(10):1219–1222, 1983.

[97] S. Osowski and T.H. Linh. ECG beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering*, 48(11):1265–1271, 2001.

[98] Y. Ozbay, R. Ceylan, and B. Karlik. A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine*, 36(4):376–388, 2006.

[99] J. Pan and W.J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, pages 230–236, 1985.

[100] K. Porter, J. Ahlgren, J. Stanley, and L.F. Hayward. Modulation of heart rate variability during severe hemorrhage at different rates in conscious rats. *Autonomic Neuroscience: Basic and Clinical*, 150(1-2):53–61, 2009.

[101] T.J. Regan, M.A. Harman, P.H. Lehan, W.M. Burke, and H.A. Oldewurtel. Ventricular arrhythmias and K+ transfer during myocardial ischemia and intervention with procaine amide, insulin, or glucose solution. *Journal of Clinical Investigation*, 46(10):1657, 1967.

[102] J.K. Romberg, H. Choi, R.G. Baraniuk, and N. Kingsbury. A hidden Markov tree model for the complex wavelet transform. *IEEE Trans. Signal Processing*, 2002.

[103] W. Rutland-Brown, J.A. Langlois, K.E. Thomas, and Y.L. Xi. Incidence of traumatic brain injury in the United States, 2003. *The Journal of Head Trauma Rehabilitation*, 21(6):544, 2006.

[104] E. Sandøe and B. Sigurd. *Arrhythmia: a guide to clinical electrocardiology*. Publishing Partners Verlags, 1991.

[105] Kellermann A Lormand JD. Sasser S, Varghese M. Prehospital trauma care systems. *Geneva: World Health Organisation*, 2005.

[106] A. Sauaia, F.A. Moore, E.E. Moore, J.B. Haenel, R.A. Read, and D.C. Lezotte. Early predictors of postinjury multiple organ failure. *Archives of Surgery*, 129(1):39, 1994.

[107] A. Sauaia, F.A. Moore, E.E. Moore, K.S. Moser, R. Brennan, R.A. Read, and P.T. Pons. Epidemiology of trauma deaths: a reassessment. *The Journal of trauma*, 38(2):185, 1995.

[108] H. Schmal, M. Markmiller, A.T. Mehlhorn, and N.P. Sudkamp. Epidemiology and outcome of complex pelvic injury. *Acta Orthopedica Belgica*, 71(1), 2005.

[109] A. Schuck Jr and JO Wisbeck. QRS detector pre-processing using the complex wavelet transform. In *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, volume 3, 2003.

[110] PJ Schwartz and SG Priori. Sympathetic nervous system and cardiac arrhythmias. *Cardiac Electrophysiology: From Cell to Bedside. Philadelphia, Pa: WB Saunders Co*, pages 330–343, 1990.

[111] PJ Schwartz and S. Wolf. QT interval prolongation as predictor of sudden death in patients with myocardial infarction. *Circulation*, 57(6):1074, 1978.

[112] A. Scope, U. Farkash, M. Lynn, A. Abargel, and A. Eldad. Mortality epidemiology in low-intensity warfare: Israel Defense Forces experience. *Injury*, 32(1):1–3, 2001.

[113] I.W. Selesnick, R.G. Baraniuk, and N.C. Kingsbury. The dual-tree complex wavelet transform. *Signal Processing Magazine, IEEE*, 22(6):123–151, 2005.

[114] E.B. Sgarbossa, S.L. Pinski, A. Barbagelata, D.A. Underwood, K.B. Gates, E.J. Topol, R.M. Califf, G.S. Wagner, et al. Electrocardiographic diagnosis of evolving acute myocardial infarction in the presence of left bundle-branch block. *The New England Journal of Medicine*, 334(8):481, 1996.

[115] A. Skodras, C. Christopoulos, and T. Ebrahimi. The JPEG 2000 still image compression standard. *Signal Processing Magazine, IEEE*, 18(5):36–58, 2001.

[116] M.H. Song, J. Lee, S.P. Cho, K.J. Lee, and S.K. Yoo. Support vector machine based arrhythmia classification using reduced features. *International Journal of Control Automation and Systems*, 3(4):571, 2005.

[117] Y. Sun, K.L. Chan, and S.M. Krishnan. ECG signal conditioning by morphological filtering. *Computers in biology and medicine*, 32(6):465–479, 2002.

[118] Y. Sun, K.L. Chan, and S.M. Krishnan. Characteristic wave detection in ECG signal using morphological transform. *BMC Cardiovascular disorders*, 5(1):28, 2005.

[119] Y. Suzuki. Self-organizing QRS-wave recognition in ECG using neural networks. *IEEE Transactions on Neural Networks*, 6(6):1469–1477, 1995.

[120] MP Tarvainen, PO Ranta-Aho, and PA Karjalainen. An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, 49(2):172–175, 2002.

[121] PE Trahanias. An approach to QRS complex detection using mathematicalmorphology. *IEEE Transactions on Biomedical Engineering*, 40(2):201–205, 1993.

[122] MG Tsipouras, DI Fotiadis, and D. Sideris. An arrhythmia classification system based on the rr-interval signal. *Artificial Intelligence in Medicine*, 33(3):237–250, 2005.

[123] E.D. Ubeyli. ECG beats classification using multiclass support vector machines with error correcting output codes. *Digital Signal Processing*, 17(3):675–684, 2007.

[124] National Highway Traffic Safety Administration US Department of Transportation. Traffic safety facts 2004. a compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system. *Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration.*

[125] RF von Borries, JH Pierluissi, and H. Nazeran. Wavelet transform-based ECG baseline drift removal for body surface potential mapping. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 3891–3894, 2005.

[126] A.L. Wade, J.L. Dye, C.R. Mohrle, and M.R. Galarneau. Head, face, and neck injuries during Operation Iraqi Freedom II: results from the US navy-marine corps combat trauma registry. *The Journal of Trauma*, 63(4):836, 2007.

[127] E.P. Walsh and JP Saul. Cardiac arrhythmias. *Nadas Pediatric Cardiology. Philadelphia, Hanley & Belfus Inc/Mosby*, pages 377–433, 1992.

[128] I. Wilcox and C. Semsarian. Obstructive Sleep Apnea: A Respiratory Syndrome With Protean Cardiovascular Manifestations. *Journal of the American College of Cardiology*, 54(19):1810, 2009.

[129] J.F. Williams Jr, D.L. Boyd, and J.F. Border. Effect of acute hypoxia and hypercapnic acidosis on the development of acetylstrophanthidin-induced arrhythmias. *Journal of Clinical Investigation*, 47(8):1885, 1968.

[130] G. Wu and E.Y. Chang. Class-boundary alignment for imbalanced dataset learning. In *Proceedings of the ICML*, volume 3. Citeseer, 2003.

[131] Q. Xue, YH Hu, and WJ Tompkins. Neural-network-based adaptive matched filtering for QRS detection. *IEEE Transactions on Biomedical Engineering*, 39(4):317–329, 1992.

[132] W. Yu, T. Liu, R. Valdez, M. Gwinn, and M.J. Khoury. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making*, 10(1):16, 2010.

[133] F. Zhang and Y. Lian. Electrocardiogram QRS detection using multiscale filtering based on mathematical morphology. pages 3196–3199, 2007.

[134] F. Zhang and Y. Lian. QRS Detection Based on Multiscale Mathematical Morphology for Wearable ECG Devices in Body Area Networks. *IEEE Transactions on Biomedical Circuits and Systems*, 3(4):220–228, 2009.

VITA

Abed Al Raoof Bsoul received a BSc. degree in computer science from Yarmouk University, Irbed, Jordan in 2002. Two years later, he got his master degree in computer science and information from the same university. His research interests include biomedical signal and image processing, medical informatics decision making and information retrieval.

List of Relevant Publications:

Yurong Luo, **A. A. R. Bsoul**, Kayvan Najarian, "Confidence Based Classification with Dynamic Conformal Predication and Its Applications in Biomedicine", 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11).

A. A. R. Bsoul, Soo-Yeon Ji, Kevin Ward, Kayvan Najarian, "Automatic Prediction of Arrhythmia Severity Using Time and Frequency Domain Features", American Heart Association, Circulation 2010;122:A213.

Soo-Yeon Ji, Kevin Ward, **A. A. R. Bsoul**, Yurong Luo, Kathy Ryan, Caroline Rickards, Victor Convertino, Kayvan Najarian, "Developing a Hypovolemia Monitoring System by Integrating Physiological Measures and P-QRS-T Waves", American Heart Association, Circulation 2010;122:A272.

Fadi Obeidat, Jose Ortiz, Jeremy Cooper, **Bsoul A.A.R.**, and Robert Klenke "Embedded Systems Performance Modeling using FPGA-based Profiling", In Proc. of The 2010 International

Conference on Embedded Systems and Applications (ESA'10), Las Vegas, July-2010

Bsoul A.A.R., S. Ji., K. Ward, and K. Najarian, Detection of P, QRS, and T Components of ECG using Wavelet Transformation," : International Conference on Complex Medical Engineering (CME 2009).

Bsoul, A.A.R. and Ji, S.Y. and Ward, K. and Ryan, K. and Rickard, C. and Convertino, V. and Najarian, K. Prediction of Severity of Blood Volume Loss Using ECG Features Based on P, QRS, and T Waves , American Heart Association, Circulation. 2009; 120:S1466.

Ji, S.Y., **Bsoul, A.A.R.**, Ward, K., Ryan, K., Rickard, C., Convertino, V. and Najarian, K. Incorporating Physiological Signals to Blood Loss Prediction Based on Discrete Wavelet Transformation, American Heart Association Circulation. 2009; 120:S1483.

List of Relevant Patents:

Bsoul A. A. R., S. Ji, K. Ward, and K. Najarian, "Detection and Classification of Arrhythmia Severity using Support Vector Machine and Deterministic Finite Automate", Filed 2011.

Bsoul A. A. R., S. Ji, K. Ward, and K. Najarian, "Methods and aparatus for determining heart rate variablitity using wavelet transform", Filed 2010.

Bsoul A. A. R., S. Ji, K. Ward, and K. Najarian, "Prediction of critical injuries and illnesses using real-time signal and analysis of multi physiological signals collected by mobile physiological monitors", Filed 2010.

Bsoul A. A. R., S. Ji, K. Ward, and K. Najarian, ”Detection of P, QRS, and T components of ECG using Wavelet Transformation”, Filed 2009.