HEARTWAVE BIOMETRIC AUTHENTICATION USING MACHINE LEARNING ALGORITHMS

Chin Leng Peter, LIM

A thesis submitted to the Newcastle University for the degree of

Doctor of Philosophy

School of Engineering

Faculty of Science, Agriculture and Engineering

December 2018

ABSTRACT

The advancement of IoT, cloud services and technologies have prompted heighten IT access security. Many products and solutions have implemented biometric solution to address the security concern. Heartwave as biometric mode offers the potential due to the inability to falsify the signal and ease of signal acquisition from fingers. However the highly variated heartrate signal, due to heartrate has imposed much headwinds in the development of heartwave based biometric authentications.

The thesis first review the state-of-the-arts in the domains of heartwave segmentation and feature extraction, and identifying discriminating features and classifications. In particular this thesis proposed a methodology of Discrete Wavelet Transformation integrated with heartrate dependent parameters to extract discriminating features reliably and accurately.

In addition, statistical methodology using Gaussian Mixture Model-Hidden Markov Model integrated with user specific threshold and heartrate have been proposed and developed to provide classification of individual under varying heartrates. This investigation has led to the understanding that individual discriminating feature is a variable against heartrate.

Similarly, the neural network based methodology leverages on ensemble-Deep Belief Network (DBN) with stacked DBN coded using Multiview Spectral Embedding has been explored and achieved good performance in classification. Importantly, the amount of data required for training is significantly reduced.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisors: Dr Wai Lok Woo and Prof Satnam Dlay for their dedications and professionalism which have guided me tremendously in my research work. They have been very kind, patience and tolerant to my slow pace and constantly encouraging me to explore unchartered domains. Without them, this dissertation and research would not have been possible.

I would like to take this opportunity to express my gratitude to my organization and my colleagues for their strong. Their support have allowed me to focus on my thesis without any disruption from work.

Most importantly, I owe my almost impossible task of PhD completion to my family, my kids (Rachel and Royden) and in particular to my beloved wife, Sally. The kids have been a significant emotional and mental encouragement for me. The relentless love and care from Sally and as guardian of the family, have enabled me to pursue this PhD without any burden or beholden.

Lastly but not least, the Almighty God for the wisdom and knowledge that enabled me to persevere throughout the stressful periods.

LIST OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS/ACRONYMS

bpm	Beats Per Minute
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
PVC	Premature Ventricular Contraction
HMM	Hidden Markov Model
GMM	Gaussian Mixture Model
GMM-HMM	Gaussian Mixture Model-Hidden Markov Model
DBN	Deep Belief Network
MSE	Multiview Spectra Embedding
AC/DCT	Auto Correlation with Discrete Cosine Transformation
AC	Auto Correlation
DCT	Discrete Cosine Transform
AR	Auto Regression
HRV	Heart Rate Variability
ECG	Electrocardiogram
SVT	Super Ventricular Tachycardia
NN	Neural Network
ANN	Artificial Neural Network
CNN	Convolution Neural Network
MCNN	Multiresolution Convolution Neural Network
HDP-AR-HMM	Hierarchical Dirichlet Auto-regression Hidden Markov Model
DWT	Discrete Wavelet Transformation
MRA	Multi-Resolution Analysis

QMR	Quadrature Mirror Filters
DB4	Daubechies 4th Order Wavelet
aVR	Augmented Vector Right
PPV	Positive Predictive Value
EER	Equal Error Rate
EM	Expectation Maximization
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MDL	Minimum Description Length
FN	False Negative
FP	False Positive
GBRBM	Gaussian-Bernoulli Restricted Boltzmann Machine
TP	True Positive
TN	True Negative
RBM	Restricted Boltzmann Machine
ELM	Extreme Learning Machine
SVM	Support Vector Machine
CD	Contrastive Divergence
SNR	Signal-Noise-Ratio

LIST OF PUBLICATIONS

[1] C. L. P. Lim, W. L. Woo, and S. S. Dlay, “Enhanced Wavelet Transformation for Feature Extraction in Highly Variated ECG Signal” *(IET Conference Proceedings). Institution of Engineering and Technology, 2015, pp. 6 .-6*

[2] C. L. P. Lim, W. L. Woo, S. S. Dlay, B. Gao and A. Bouridane, “Heartwave Features Sensing of Time-Varying QT and PR Interval under Extreme Exercise Duress”, *submitted to IEEE Transactions on Instrumentation and Measurement.*

[3] C. L. P. Lim, W. L. Woo, S. S. Dlay and Bin Gao, “Heartrate-Dependent Heartwave Biometric Identification with Thresholding-Based GMM-HMM Methodology”, *In-Press, IEEE Transactions on Industrial Informatics.*

[4] C. L. P. Lim, W. L. Woo, S. S. Dlay, Di Wu and Bin Gao, “Deep Multi-View Heartwave Authentication”, *In-Press, IEEE Transactions on Industrial Informatics.*

[5] C. L. P. Lim, W. L. Woo and S. S. Dlay, “An Investigation on Intra-User and Inter-User ECG Based Biometric with Generative Modelling under Variable Heartrate”, *accepted* , *2018 10th International Conference on Machine Learning and Computing (*ICMLC 2018*).*

CHAPTER 1

INTRODUCTION TO HEARTWAVE BASED BIOMETRIC CLASSIFCATION

This chapter introduces the advantages and the needs of using heartwave signal as a biometric mode. Conversely, it highlights the difficulty in using heartwave signal as a biometric mode in particular to elevating heartrate. The chapter will elaborate on the states of heartwave components: P-Wave, QRS-Complex and T-Wave under varying heartrates. Chapter 1 will also establish the objectives of this research and a brief overview of each chapter.

1.1 Heartwave based Biometric Authentication

1.1.1 Potential Heartwave based Biometric Application

The use of individual heartwave signal as biometric mode to perform identification present many opportunities. With increasing evolution of digital technology, IoT solutions, cloud services and big data services, the need for secured data protection is univocal and many have implemented 2-Factor Authentication (2FA) similar to the security adopted by ebanking.

In E-health management, heartwave as a biometric mode has great potential to fulfil the security demands and ensuring access integrity [1]. With increasing elderly population and longer life expectancy, elderly suffering from worn-out fingerprint and poor eyesight are facing challenges to use security system such as Digital Key Token and fingerprint biometric system comfortably. Equally, it is tormenting for elderly to setup password with periodic renewal and adhering to unique password characters combination. Heartwave as biometric mode has great potential to complement existing 2FA infrastructure for secured access to services and products through the means of wearable devices embedded with electrodes for heartwave signal acquisition. In addition, it eliminates risks involving retrieval of wrong medical records due to similar names.

In transportation, heartwave signal as biometric mode can be used to enhance transportation safety such as authenticated access to vehicle with continuous monitoring of driver fatigue due to prolong driving [2, 3].

In healthcare, there are intense developments in tele-health solutions to provide continuously monitoring on the well-being of the elderly [3]. Biometric authentication for access to services enables medical personnel to respond to elderly needs reliably, securely and promptly. See Figure 1.1 for illustration.

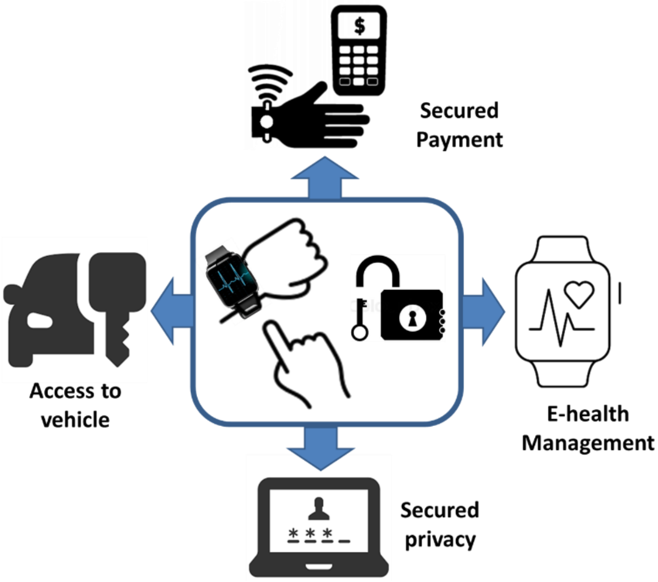


Figure 1.1: Applications of heartwave as biometric mode

1.1.2 Review of Current Biometric Modes

Reviews of various biometric techniques that have been explored [4-10]. Biometric can be broadly classified under biological biometric and behavioral biometric.

Behavioral biometrics deals with the patterns and gaits of human movement that includes, voice, handwriting and signature. Biological biometric deals with individual anatomy and detailed physiology.

This review examines the characteristic of the current biometric modes, the disadvantages and advantages relative to this research tenet of heartwave signal as biometric mode. Table 1.1 summarizes the characteristics of the reviewed biometric modes.

Table 1.1: Comparison of biometric modes

Biometric Mode	**Accept Ratio**	**Reject Ratio**	**Accuracy**	**Cost**	**Device Required**
Iris	66%	34%	High	High	Camera
Retina	75%	25%	High	High	Camera
Vein	92%	8%	High	Medium	Camera
Facial	85%	15%	Medium-Low	Medium	Camera
Fingerprint	80%	20%	High	Medium	Scanner
Heartwave	95%	5%	High	Low	Electrodes [11]

1.1.2.1 Biometric Mode: Face

In face recognition, features consisting of eyes, eyebrow, noise, mouth and inter-feature relatives distances are used as the discriminating factors in authentication. Approaches used in facial recognition include Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The implementation of facial recognition requires camera vision to capture the facial of individual. Extraction requirements also include the ability to extract foreground feature from background image. As image recognition is employed, the acquisition process can be affected by the lighting condition, spectacles, ageing effects, facial expression and position of the user during acquisition. In the early formative years, individual facial experiences craniofacial growth where face becomes larger. In the transition to elderly, individual facial experiences deformation of the skin such as texture, thinner, darker, less elastic, more wrinkles and blemishes. However texture transformation information are not use as discriminative feature. Conversely, minus the mentioned constraints, facial recognition is non-invasive and can be easily implemented due to low capital requirement and easily available implementable code.

1.1.2.2 Biometric Mode: Fingerprint

Fingerprint as biometric mode has been most widely used [12]. Fingerprint is a unique characteristic traits that exist in all individual including twins. Fingerprint contains patterns of ridges and valleys that form the uniqueness among individuals. The unique traits are also known as minutiae. Typically a fingerprint consists of forty minutiae and similar minutiae between two individuals are less than eight.

This technology is most widely adopted, due to the availability of fingerprint sensor at an extremely affordable price. In addition, code to develop a fingerprint authentication module are readily shared at community level. Fingerprint biometric mode however has seen reports of attempt falsifications [13] [14]. More apparently, fingerprint is not permanence as the size and shape changes as the finger grows. In addition, fingerprint degrades over time. Elderly has difficulty in using fingerprint as the biometric mode due to worn out ridges on the fingers. Individual with sweaty palms suffered from using fingerprint biometric system efficiently. Hence, in elderly and individual with sweaty palms, using fingerprint biometric system is more of the chore rather than an ease of access to them.

1.1.2.3 Biometric Mode: Iris

In iris biometric system, the recognition consists of acquiring the image of the iris for authentication. Iris is a muscular tissue that regulates the size of the pupil. The formation of the muscle is structurally distinct that allows for authentication purposes.

Iris pattern is unique between individuals and the pattern achieves permanence characteristic before the age of one which is a highly desirable biometric trait. One of the challenges of iris as the biometric mode is image acquisition. The ability to authenticate an individual accurately is highly dependent on the image quality without causing discomfort to users. Accurate it may be, iris imaging can be affected by eyelashes and reflection. More worrying, its uses infrared light to illuminate the iris which many view as invasive.

1.1.2.4 Biometric Mode: Retina

Similar to iris, retina is another biometric mode. In retina scanning, the network of blood vessel is use as a unique characteristic traits to authenticate user. It has been reported the network of blood vessel is unique even in twins. In the acquisition of the retina for authentication, the acquisition system has to be placed five to eight centimeters from the eyes for the image to be captured reliably. More importantly, retina scanning uses infrared signal that is projected into the eye where it is absorbed by the blood vessel in the retina. This absorption causes illumination of the retina and which can be captured for authentication. As the mode of acquisition uses infrared, it is conveniently classified as an invasive authentication. Similarly to iris based authentication, accuracy highly dependent on the image quality without causing discomfort to users. Although extremely accurate, the method can be affect by individual with eye related medical conditions and diseases such as cataract and glaucoma.

1.1.2.5 Biometric Mode: Palm Veins

In palm vein biometric mode, it relies on the network of blood vessels on the palm as a discriminating feature. The network of blood vessel is acquired by passing near infrared light through the body and the reflected image is captured. Results are highly accurate, unlike fingerprint, it is not affected by dry or sweaty palm. However, the equipment required to capture the palm vein is bulky and not suitable to be implemented in wearable device.

1.1.2.6 Heartwave as Biometric Mode

Heartwave is a biological signal generated by the rhythmic electrical activity of the heart. It also represents the liveliness of an individual. Apart from possessing the essential biometric characteristic traits, the impossibility to falsify and mimic the biological signal makes heartwave signal an ideal candidate as the biometric mode [15, 16]. Concurrently, heartwave signal can also be used to provide addition information pertaining to psychological, physiological and clinical status of an individual [17].

The uniqueness of heartwave signal is supported by 2 primary factors: physiological (structural construction of heart) condition of the heart and the geometrical (chest geometry) location of the heart. The two primary factors manifested a heartwave signal that is unique between individuals [18].

Many attempts have been made to introduce heartwave as a form of human identification mode due to the pertinent fact that the biological activity and geometrical features of the heart have resulted a unique signature in all individuals. The challenges faced are wide and varying. Heartwave based biometric is a multi-stage process that involves signal acquisition from individual under non-clinical setup condition, signal conditioning where interfering noises are eliminated or reduced while maintaining data integrity, extraction of key distinctive data that are repeatable yet unique for discriminating process, recognition and matching where high dimensional data are analyzed, matched and making informed decision to either grant or deny entry to an individual[12, 18].

Individual heartrate is variable and the impetus of variations can be contributed by many factors such physiological activities, psychological related and pathological related issues. The heartrate variation of individual can vary as much as 400%. Even in resting, variation of the heartwave signal exists due to movement of the respiratory cage [19, 20]. Heartwave morphology which is dependent on heart rate, suffers from minimal to signification variations. Every individual has its own resting heartrate and maximum heartwave. At resting state, heart rate variation is at minimal and the heart rate of an individual can range from 50 bpm to as much as 180 bpm in accordance to the maximum heart rate equation of “220 bpm – age of an individual”. Hence, since heartwave morphology varies according to heartrate and as a biometric mode, reliable extraction of heartwave features is utmost critical to enable reliable identification and verification.

The use of heartwave signal as biometric mode has aroused many research works with approaches such as KNN classifiers [18, 21, 22], LDA classifier [23], Support Vector Machine, Match Score Classifier [24] and Generative Model Classifier [25-27]. Unfortunately, all of the above works use ECG data that are obtained under resting condition where individual heartrate is not under physical duress. As mentioned, the morphology of an individual heartwave changes under different heartrate. One reported work [28] uses data comprises of heartwave signal under varied conditions of heart wellness. The work uses auto-correlation method to discard anomaly waveform of Premature Ventricular Contraction (PVC). PVC is a heart anomaly signal that occurs sporadically unlike the repetitive heartwave signal. Linear Discriminant Analysis is subsequently used to perform classification. Although categorized under varied conditions, the work does not use signals that are acquired under physical duress. In medical related fields, the works [25, 27, 29] use signal processing tools and Hidden Markov Model (HMM) to detect heartwave anomaly on individual with cardiac related problem. In those works, heartwaves of multiple individuals of different anomalies are concatenated and input to HMM model for anomaly detection. The HMM model is thus not appropriate for individual classification since individual heartwave morphologies varies across individual heart rate.

1.2 Heartwave Signal and the Effects of Varying Heartrate

Heartwave signal comprises of 3 main wave complexes that are important in heartwave base authentication and identification. The three waves are namely P-Wave, QRS-Complex and T-Wave. In a single heartwave signal, it starts off with the contraction of the atrium muscle tissue which results in the formation of the P-Wave. Upon contraction, the excited electrical pulse travels to the ventricular muscle causing contraction which produces the QRS-Wave complex. QRS-Wave complex is the most recognizable peak and is caused by the large muscular tissue mass at the ventricular segment. Following the ventricular contraction is relaxation of the ventricular which causes the formation of the T-Wave. See Figure 1.2 for illustration.

Though rhythmic and continuous in nature, the depolarization of the atrial is known as the start of the cardiac cycle.

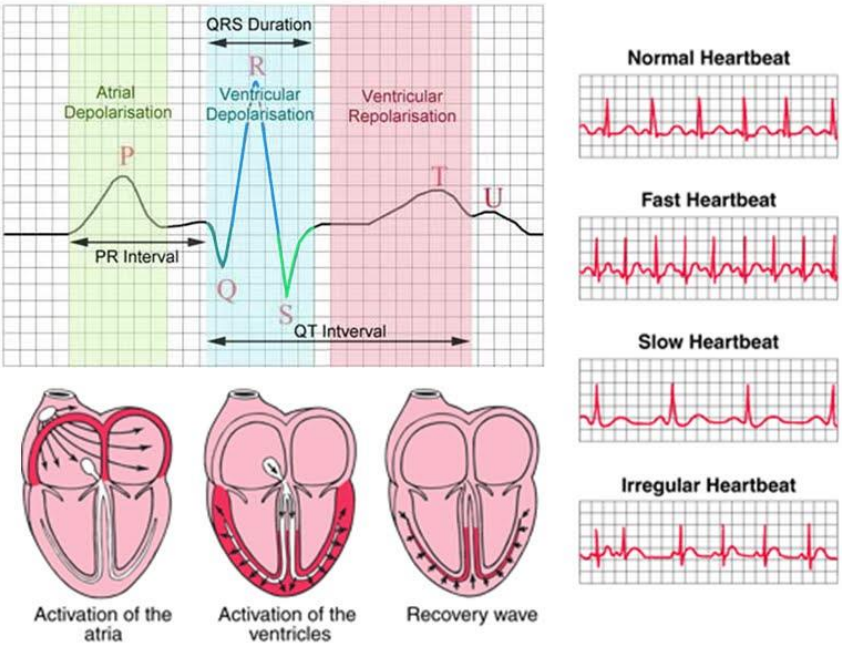


Figure 1.2: Traces of heartwave signal

Medically, it has been proven that individual heartwave experiences persistent variations. Numerous studies have been done to conclude that heartwave morphology varies regardless of individual state of health, gender sex and age. In the medical practice, individual heartwave signals are analyzed in a multi-stage exercise also known as ECG treadmill where individual will be subject to increasing exercise duress till the individual maximum heart rate is breached. The analysis includes the direction and magnitudes of heartwave components consisting of P-Wave, QRS-Complex and T-Wave. Heartwave morphological variations cannot be understated [30-36]. Figure 1.3 shows the variation of heartwave at different stages of the heartrate in particular to the variations of T-Wave. At resting heartrate, it exhibits an inverted T-Wave. However at elevated heartrate, T-Wave actually slops upward into the positive region. Incidentally, it also illustrates an almost conjoint P-Wave and T-Wave.

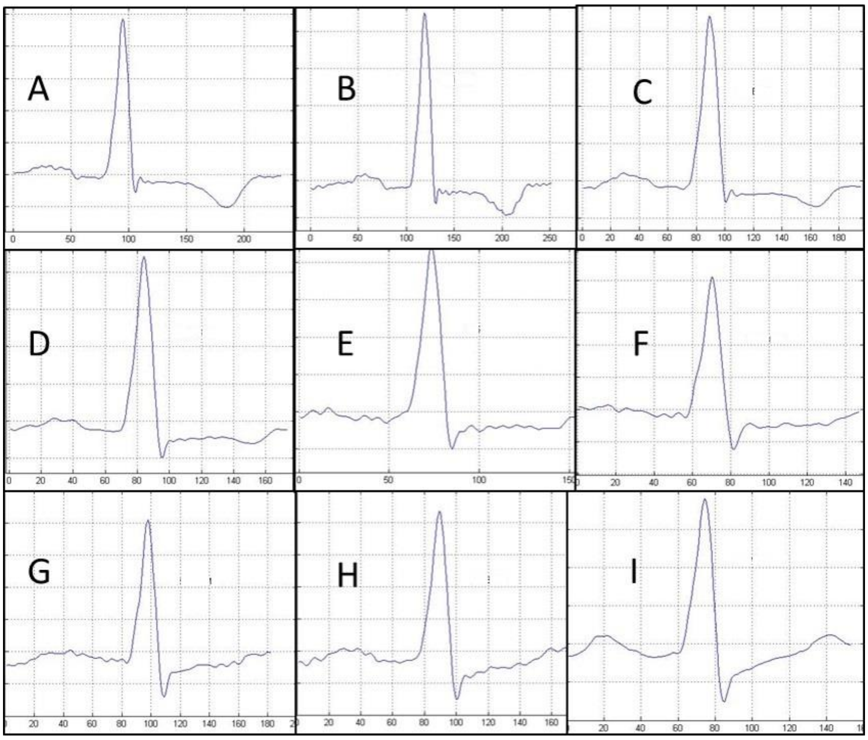


Figure 1.3: Changes of an individual heartwave morphology variations under different heart rates from a resting heart rate of 63 bpm: a. 77 bpm at 120%, b. 88 bpm at 140%, c. 100 bpm at 160%, d. 113 bpm at 180%, e. 126 bpm at 200%, f. 138 bpm at 220%, g. 150 bpm at 240%, h. 164 bpm at 260% and i. 177 bpm at 280%

1.2.1 P-Wave

P-Wave typically refers to the PR interval that spans from the onset of P-Wave characteristic point to the peak of R-Wave characteristic point. Medically, PR interval is the time taken from the activation of the atrium to the Purkinje Conduction System by the transmission of electrical stimulation. PR interval is reversely and linearly correlated to the heart rate that ranges from 55bpm to 155bpm. During strenuous exercise, P-Wave morphology changes and is inversely proportional to the change in heartrate in which the PR interval shorten with increasing magnitude [37] [38]. In summary, the interval between the spatial maximum between P-Wave and the onset of the QRS-Complex of the P-Wave decreases while the magnitude of the P-Wave increased. It is also to note that while the interval and magnitude change, the direction of the P-Wave vector remains unchanged. Using a mean vector to quantify the P-Wave, the P-Wave vector magnitude increases approximately by 68% [38]. In another studies, the PR interval can be quantified at 5 milliseconds reduction for every increases of 10 heart rates per minute [39].

In addition to heartwave morphology variation during exercise, heart activity undergoes a recovery period immediately after an intense exercise. During the recovery period, P-Wave magnitude increases remarkably during the first minute of the recovery with lengthening of PR temporal period before recovering to resting state condition [39] [33] [37]. Currently none has studied into P wave uniqueness to support authentication. No study has been conducted on the effects of P wave variation on heartwave based biometric authentication [24].

1.2.2 QRS-Complex

QRS-Complex component is most distinctive segment of the heartwave signal as it contains the most prominent R-Wave peak. QRS-Complex is relatively stable with minimal variation during resting condition. During strenuous exercise, the QRS period is shorten similarly to P-Wave. However the variation occurs in the segment of RS while segment of QR exhibits minimal variation during exercise [32]. In addition to the temporal variation, magnitude of the QRS complex also varies. R-Wave actually decreases in magnitude while S-Peak increases. In some situation, the S-Peak can be higher than R-Peak [33] [30].While QRS-Complex exhibits unique variation during exercise, the unique variation is also observe during the recovery stage. During recovery, QR interval actually lengthen during recovery. The variation ceases and morphology stabilizes once heartrate returns to resting condition [33].

1.2.3 T-Wave

T-Wave exhibits the most significant variation. Medically, it is known as the QT interval. The QT interval initiates from the onset of Q-Wave to the offset of T-Wave. Medically, QT interval has been used to deduce pathological causes of heart related ailments and widely tracked to determine the efficacy of cardiac related drugs [40] [41] [42].

Signal profile is highly dependent on the Lead position where the signal is acquired. Signal acquired from Lead III often exhibits signal in negative domain. During exercise, the QT interval will shorten with decrease in T-Wave magnitude due to increase heart rate. In short, the QT interval is reversely linear to the change in heart rate [36]. During recovery from exercise, the QT interval will lengthen due to decrease of the heart rate [33] [37]. A unique observation is the spike of the T-Wave peak during the first minute of the recovery [33]. Hence while the QT period is reversely linear to change in heart rate, the vector change is independent of heart rate [37].

As mentioned on the use of QT interval as a prognosis and diagnosis for heart anomaly, various methods such as Framingham Method have been developed to correct QT variations and the signal is reconstructed in 60bpm interval [43, 44] [45]. However the correction technique does not address the spatial magnitude variation.

1.3 Aim of Thesis

The objective of this research focuses on investigating and developing a machine learning methodology and architecture to address individual heartwave morphological variations and performing a biometric classification that is invariant of heartrate and heartwave variations. The variations of heartwave signal contributed by physical activities and varying heartrates have caused current reported authentication and classification methodologies to deteriorate significantly in intra-person and inter-person authentication.

The objectives of the research are as follows:

1. To present a unified perspective of the commonly adopted methods to support heartwave segmentation and extraction, and methods of determining discriminating feature and classification.

2. To develop a novel approach that is capable to perform heartwave segmentation and feature extraction reliably and accurately under variable varying heartrate conditions.

3. To develop a statistical modelling architecture that is able to identify individual unique and varying discriminating features due to varying heartrate and perform individual classification accurately and reliably based on the developed model.

4. To develop a deep neural network based architecture that is able to harness and accommodate individual varying heartwave signal and perform feature extraction to produces a high performance biometric classification methodology.

5. To analyze the proposed approaches and evaluate the classification performance with other state-of-the-art approaches.

1.4 Thesis Outline

This thesis focuses on two aspects. The first aspects focuses on the heartwave segmentation and extraction to extract heartwave characteristic features under extreme heartwave morphological variations. The second aspect focuses on discriminative data clustering and classification using two different domains: generative modeling methodology and deep neural network methodology. The theoretical concepts of each algorithm and architecture will be introduced in details. All the experiments, simulations and validations are conducted on Matlab using Databases from Physionet. Results are compared with state-of-the-art methodology and discussed. The outline of the thesis is as follows:

Chapter 2 provides a comprehensive review of the current state-of-the-art methodologies in the domains of heartwave signal segmentation and extraction, and the classification. Each of the methods is described and the appropriateness to support the scopes of the research will be discussed. Reported works based on the methodology will be represented with the merits and issues discussed.

Chapter 3 focuses on the proposed novel methodology of Discrete Waveform Transformation integrated with heartrate dependent parameters: PR-Interval and QT-Interval to perform specific waves' components delineation and extraction. The proposed methodology is able to perform segmentation and extraction of heartwave features regardless of heartrate.

Chapter 4 features a novel statistical methodology that leverages on the characteristic strength of Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM). The proposed architecture comprises of GMM-HMM coupled with user specific parameters of thresholding criteria from loglikelihood score and heartrate are used to perform user classification. Each of the user dataset is modelled by GMM and subsequent remodel under HMM to provide discriminating features for individual classification. As heartwave morphology is closely tied to the heartrate and each user has a determined range of heartrate, these parameters become unique features for classification.

Chapter 5 features a novel architecture that uses an ensemble of Deep Belief Networks (DBN) to perform heartwave extraction under different configurations. Thereafter, a method of Multiview Spectra Embedding (MSE) is adopted to combine the output of the ensemble-DBN into a single structure of data. Importantly, the MSE identify key significant feature from each of the views and combined the views as complimentary property. The output of the MSE is further input into a stack DBN to perform classification of individual.

Chapter 6 concludes the research work of this thesis followed by recommendation of future works that are necessary to conceptualize the heartwave based classification for future deployment.

CHAPTER 2

OVERVIEW OF HEARTWAVE BASED AUTHENTICATION

2.1 Introduction

In this chapter, the current state of the art in heartwave based authentication is introduced in details. In heartwave based biometric authentication, the solution typically comprises of two domains. The domains are heartwave segmentation with feature extraction and classification techniques. The two domains are important and critical. A low quality segmentation and extraction process have the consequences of affecting the classification outcome. This chapter comprises of three sections. Section 1 presents the current state-of-the-art solution to perform classification under cardiac irregularity and under exercise duress. Section 2 elaborates on the methodologies that have been reported to support heartwave signal segmentation and extraction. Section 3 elaborates on the current methodologies to support feature identification and classification.

2.2 Heartwave Biometric Solution for Cardia Irregularities and under Exercise Duress

Reported work on the heartwave biometric solution for cardiac irregularities and under exercise duress are far and few. To-date, to the best of author’s knowledge, work in the mentioned context is only a hand full. There are unfortunately, more work on identifying and discriminating different variants of anomaly heartwave signal to support medical intervention. [22, 46-51] have reported work on the biometric classification involving individuals with anomaly heartwave signal.

[52] reported a work on heartwave based identification on individuals with normal and cardiac anomalies. The work uses analytical approach to segment the peaks and valleys around R-Peak. The window for peaks and valleys detection is limited at 25-sample length about R-Peak. Thereafter, a commercial software Kubios is used to remove artifact signal. Random Forest methodology is used to determine discriminating feature and perform classification. The work is performed on the databases containing normal signal and arrhythmia signal and reported an accuracy of 95.85%. While commendable, a few key essentials are not covered. For each record, only 25 heartbeat samples are used for testing which suggests that significant amount of data has been used for training. Secondly, although testing has been conducted on arrhythmia database with anomaly heartwave, the anomaly heartwave signal is under resting state. Thirdly, with a fixed window of 25-sample for peak and valley detection, it is not possible to detect heartwave at elevated heartrate. At elevated heartrate, the profile of the T-Wave has a tendency to exhibit an upslope morphology to render the peak and valley detection function inappropriate.

In the work of [53], classification testing is performed on database that contains arrhythmia heartwave signal. The classification focus on QRS-Complex as the discriminating features. In the 44 individuals, there are 26 individuals with arrhythmia with sporadic occurrence of anomaly QRS-Complex. Upon extraction, the QRS-Complex is used as a template. Each healthy individual has 8 templates and 16 templates for individuals with anomaly QRS-Complex. Correlation Coefficient is used as the criteria to classify individual. Author reported an accuracy of 100%. It is however to note although testing has been conducted on arrhythmia database, the author did not investigate heartwave morphological variation under elevated heartrate. The heartwave morphological variation is highly dependent on heartrate, which leads to a contentious agreement on the number of templates that are required for each individual.

In the rare work of [54], the effect of heartwave signal variation under exercise duress is investigated. It is reported that 26 individual signals are acquired from a bicycle exercise for a period of 5 minutes and the intensity of the exercise it capped at 35% of the ***Target Heart Rate (THR)*** where it is defined as ***(Peak HR – Resting HR) x intensity% + Resting HR. Peak HR is defined as (220-age)***. The work uses chaotic theory methodology in the extraction of heartwave feature. To support classification, nonlinear Support Vector Machine with polynomial kernel is used to perform user classification. This work achieved a classification accuracy of approximately 81%. The work is commendable but it is unfortunate that the work is not conducted under the full spectrum of heartwave morphological changes. Under extreme physical duress, the intensity of the heartrate can double and triple from resting heartrate.

2.3 Heartwave Segmentation and Feature Extraction Techniques

Heartwave segmentation and feature extraction refers to the delineation of heartwave signal and the extraction of the characteristic features that can be used to perform classification and authentication. The features to be extracted can be broadly grouped into two categories. The two categories are fiducial based and non-fiducial based. In fiducial extraction, characteristic points are located on the profile of the heartwave signals. The characteristic points can include components of P-Wave, QRS-Wave Complex, T-Wave, time difference between various characteristics points and R-R interval. The components refer to the local maxima or local minima. Non-fiducial features include heartbeat segmentation to extract discriminative information without the need to determine the fiducial points. Generally, non-fiducial feature method uses the global pattern that comprises of a finite length of signal with a few heartwave signal rather than a single heartwave signal.

Over the years, various methods and techniques have been introduced to perform heartwave segmentation and feature extraction. Broadly, the techniques can be categorized into Wavelet Transformation, Auto Correlation, Chaos Theory and Auto-Regression. In addition to the mentioned methods, it is appropriate to be aware of the operating conditions of the methods which are either frequency domain, time domain or a combination of time-frequency domain. A brief description for each for the methods will be provided together with the characteristic and limitations of the methods. This is to note that the extraction techniques mentioned in the subsequent subsection are based on commonly cited approaches and by no mean exhaustive. In addition, there are also reported works that use a combination of various methods to extract fiducial and non-fiducial data for classification.

2.3.1 Wavelet Transformation

Wavelet Transformation has been the most commonly adopted technique [55-64] and has been proven to be effective against non-stationary waveform to provide good temporal localization of the heartwave components such as the onset and offset of P-Wave, QRS-Complex and T-Wave. The accuracy of the temporal localization is critical in extraction of heartwave data.

Wavelet Transform is a time scale representation and uses a linear operation that decomposes the signal into different scales (half the scale to be exact) in relation to the frequency component. The scaling identifies the low and high frequencies in the heartwave signal. In the family of wavelets used for the detection of the heartwave component, Daubechies of different orders have been chosen to perform heartwave extraction due to the close resemblance to the heartwave signal. The advantage of the DWT is the capability of reconstructing the signal from the decomposed wavelet. The Wavelet Transform of a signal is simply the summation of the signals which have been multiplied by the scaled, shifted version of wavelet function. The Wavelet Transform can be represented by

$$W\,(a,b) = \int_{-\infty}^{\infty} f(t)\Psi_{a,b}(t)dt \tag{2.1}$$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi * (\frac{t-b}{a}) \tag{2.2}$$

where $*$ represents the complex conjugation, $\Psi_{a,b}(t)$ represents the window function aka mother wavelet which in this research focus is the Daubechies wavelet. 'a' is the scale factor and 'b' is the translation factor. $(\frac{t-b}{a})$ performs a shift and scale of the mother wavelet base on dyalic grid for decomposition of the input signal. Each level of decomposition decomposes the signal into different frequency resolution and more importantly allows spatial determination of heartwave characteristic points.

[56] reported an impressive 99.8% for detection of QRS-Complex. However the results for the components of P-Wave and T-Wave are not reported. The author however did mention on imposing empirical width to contain the profile of P-Wave and T-Wave to allow peaks and valleys detection. It is to note that the recommendation is not appropriate under elevated heartrate. Under elevated heartrate, the morphological variations in particular to P-Wave and T-Wave change significantly in particular to the sloping components of the wave.

[55] uses a Continuous Wavelet Transformation and achieved an accuracy results of approximately 95% and the size of the database contains only 4 individual. Although the accuracy is impressive, the results are not conclusive due to limited users.

While Wavelet Transformation has been adopted and reported by many, it is unfortunate the proposed methodology has been tested on dataset where the heartwave signals are under resting condition. Heartwave signal morphology is highly dependent on heartrate and a fixed window, adopted by many, leads to false detection of P-Wave and T-Wave components under elevated heartrate. Chapter 3 will elaborate in details the incorporation of heartrate dependent parameters that will enable successful and reliable detection of P-Wave and T-Wave components.

2.3.2 Auto Correlation with Discrete Cosine Transformation

In addition to Wavelet Transform method, another popular method performs heartwave extraction without the requirement of fiducial point detection. [51, 65-68] have reported work on using the method of Auto Correlation with Discrete Cosine Transformation (AC/DCT) to perform heartwave extraction. The concept of AC/DCT method involves defining a window of length *N*, where the heartwave signal is trace into a non-overlapping windows which is wider than an average heartbeat length (resting heartrate). The window is of fixed length and under variable heartrates, multiple heartwave can exist inside the window. AC is applied to blend all the sequences of signal within the window through the summation of the products expressed in (2.3). With summation of product, location of the fiducials need not be explicitly determined.

$$\widehat{R_{xx}}[m] = \frac{\sum_{i=0}^{N-|m|-1} x[i]x[i+m]}{\widehat{R_{xx}}[0]} \tag{2.3}$$

where $x[i]$ is the windowed heartwave signal and $x[i+m]$ is the time shifted version of the windowed heartwave signal with a time lag of $m = 0,1,\dots,(M-1)$; $M \ll N$. M is the dimensionality of autocorrelation. The output is normalized with the $\widehat{R_{xx}}[0]$ which cancels out the biasing factor.

Since a heartwave signal contains prominent waves complexes of P-Wave, QRS-Complex and T-Wave, the wave complexes shall be the main contributors to the sum from expression (2.3) and the similar features coefficient from the moving window will be embedded. The drawback of AC is high dimensionality. Hence after the computation via AC, an energy compaction process is initiated through the application of Discrete Cosine Transform (DCT). DCT is applied to the AC coefficients for dimensionality reduction.

$$Y[u] = G[u] \sum_{i=0}^{N-1} y[i] \frac{\pi \cos(2i+u)\, u}{2N} \tag{2.4}$$

where N is the length of the signal $y[i]$ for $i = 0,1,\dots,(N - |m| - 1)$. $y[i]$ is the auto-correlated heartwave signal obtain from 2.4. $G[u]$ is expressed as the form

$$G[k] = \begin{cases} \sqrt{\frac{1}{N}}, k = 0, \\ \sqrt{\frac{2}{N}}, 1 \leq k \leq N - 1 \end{cases} \tag{2.5}$$

This allows the representation of coefficients to be presented in lower dimension and near zero component of the frequency representation to be omitted. K will contain all the significant features of the heartwave signal. Based on the reported work of [67], the features extracted using AC/DCT achieved a 94.47% accuracy in classification of users. This method however does not resolve the impact of heartrate variations. As it uses coefficients of the Auto Correlation, the coefficient will deteriorate at increasing heartrate as the heartwave waveform becomes more and more dissimilar from the resting heartwave waveform.

2.3.3 Chaos Theory Extraction

[69-72] have proposed the use of Chaos Theory to extract heartwave feature which is a non-fiducial based extraction. The concept of Chaos Theory in particular to the Lorenz Chaos System is to transform the heartwave signal into a chaos error distribution diagram which will generate a two centroid points known as chaotic eyes. This pair of chaotic eyes are subsequently use as a characteristic detection features. See Figure 2.1 for illustration. The advantage of this concept in comparison the neural network is the minimal amount of training data required for prediction.

Accordingly, the Lorenz Chaos System consists of a master (L_{master}) and a slave (L_{slave}) and are expressed in (2.6) and (2.7) respectively.

$$L_{master} = \begin{cases} x_1 = \alpha(x_2 - x_1) \\ x_2 = \beta x_1 - x_1 x_3 - x_2) \\ x_3 = x_1 x_2 - \gamma x_3) \end{cases} \tag{2.6}$$

$$L_{slave} = \begin{cases} y_1 = \alpha(y_2 - y_1) \\ y_2 = \beta y_1 - y_1 y_3 - y_2) \\ y_3 = y_1 y_2 - \gamma y_3) \end{cases} \tag{2.7}$$

Subtracting expression (2.6) and (2.7), the chaotic dynamic function of Lorenz Chaos System can be expressed as shown:

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} -\alpha & \alpha & 0 \\ \beta & -1 & 0 \\ 0 & 0 & -\gamma \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} - \begin{bmatrix} y_2 y_3 - x_2 x_3 \\ -y_1 y_3 + x_1 x_3 \\ y_1 y_2 - x_1 x_2 \end{bmatrix} \tag{2.8}$$

x is the master system with initial value of zero and y is the slave system can contains the heartwave signal values. $\alpha, \beta \; and \; \gamma$ are adjusted error coefficients which require fine tuning and optimization. $e_1 \; and \; e_2$ are used to generate the dynamical map of chaotic dynamic error. The coordinates of the two centers of gravity in the map are termed as the chaos eyes. The chaos eyes is unique and is used as a feature to support classification and identification. Therefore, each pair of eyes is unique to an individual. In the reported work of [70], the author subsequently use correlation function to perform classification of individual.

The work was tested on 35 healthy individuals and achieved an accuracy of 94.3% within a signal period of 2 seconds. Similarly, to the previous method, the work is conducted on the heartwave signal acquired under resting condition. At elevated heartrate, the features of the chaotic eyes has the tendency to become under fit to accommodate the variations of the heartwave morphological variation which leads to misclassification.

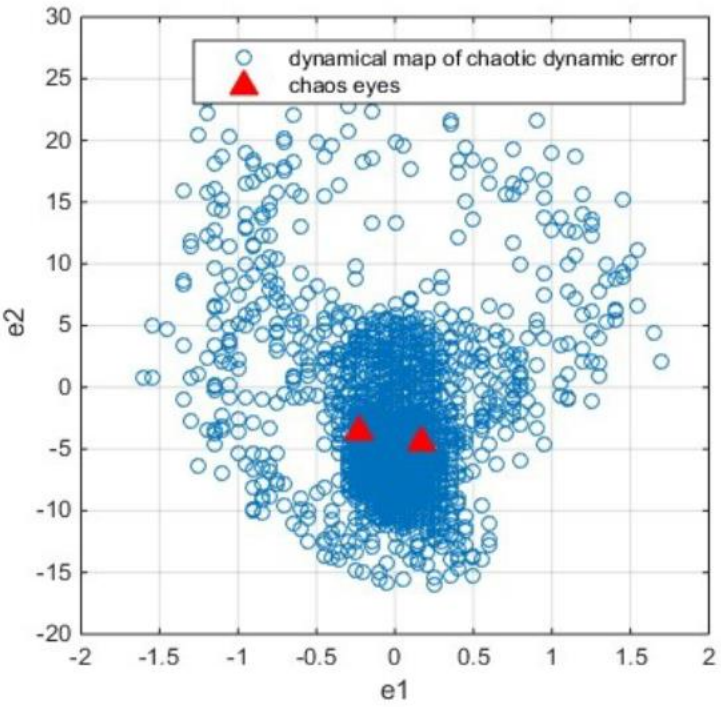


Figure 2.1: Dynamical map of chaotic dynamic error and the chaos eyes

2.3.4 Auto Regression

[73-76] has reported the use of Auto Regression (AR) to perform heartwave feature extraction. AR has been used in various applications, including classification of physiological signals like ECG, EEG, heart rate etc. The advantage of AR modeling is its simplicity and is suitable for real-time classification at the ICU or ambulatory monitoring. AR models are popular due to the linear form of the system, simultaneous equations involving the unknown AR model parameters and the availability of efficient algorithms for computing the solution. AR modeling has also been used extensively to model heart rate variability (HRV) and power spectrum estimation of ECG and HRV signals. To support heartwave feature extraction, a suitable set of features that consist of six coefficients resulting from autoregressive modeling of the three components of the ECG signal which are the P-Wave, QRS-Complex and T-Wave are used. In addition,

to enable AR method, R-Peak, being the most prominently, is separately detected. Upon detection of the R-Peak, a window of appropriate length sufficiently wide enough to accommodate a cardiac cycle is defined about R-Peak. Thus, AR is performed on a set of window length wide signal. (2.9) shows a general expression for an AR model of order P.

$$y(n) = \sum_{i=0}^{P} a_i(n)y(n-i) + e(n) \qquad (2.9)$$

where $y(n)$ represents heartwave signal, $e(n)$ represents the unknown zero mean white noise of the uncorrected random variable which is also known as modelling error. $a_i(n)$ represents the AR model coefficients. P which is the model order refers to the past samples needed to predict the present value of the data. Burg's algorithm is used to evaluate the AR coefficients. The criterion used to evaluate the model order selection is based on the sum-square error (SSE) which is defined by

$$SSE = \sum_{i=1}^{2} \sum_{j=M+1}^{N} e(j) \qquad (2.10)$$

where *N* refers to the number of sample points. A scalar AR process of order M is perform on each ECG and the each of the coefficients are concatenated to form the feature vectors for classification. It has been observed that as the value of the modeling order (M) increases, their corresponding SSE values decreases. However, for modeling orders equal or greater than three it remains constant. Consequently it can be inferred that greater values of M increase the accuracy of the classification method.

However, AR has not been report to support heartwave based biometric authentication. AR is commonly used to differentiate anomaly heartwaves which are typically, Ventricular Tachycardia, Super Ventricular Tachycardia (SVT), Premature Ventricular Contraction, Atrial Premature Contraction, Ventricular Fibrillation and the Normal Sinus Rhythm. The anomaly heartwave type are limited in classes as compared to individual classification. Hence, although promising and simplicity in development, it is not appropriate for individual classification.

2.4 Classification Techniques

Following the introduction of the current state-of-the-arts methodologies for heartwave segmentation and feature extraction, it is necessary to explore state-of-the-arts methodologies for classification. Several comparative works have been done [18, 77-79] to survey the current state-of-the-arts methodologies in classification. The commonly adopted approaches are KNN Classifier, Linear Discriminant Analysis, Neural Network Classifier, Generative Model Classifiers and Support Vector Machine. Each of the methods will be briefly presented and discussed on the appropriateness to support the heartwave classification under highly variated heartwave morphological variations.

2.4.1 KNN Classifier

K-Nearest Neighbors (KNN) is one of the simplest algorithm which is used to support statistical estimation and pattern recognition based on either similarity or distance measured. In classification, the output is a class membership. An object is classify by majority vote if its neighbors with the object being assigned to the class, is most common among its k nearest neighbor as showed in expression (2.11).

$$P(y = j|X = x) = \frac{1}{K}\sum_{i \in A} I\left(y^{i} = \mathrm{j}\right) \qquad (2.11)$$

Given a positive integer *K*, an unseen observation *x*, KNN classifier performs by running through the whole dataset, *A,* computing the distance between *x* and each training observation. The *K* in KNN refers to the number of *K* points in the training data that are closest to *x* in the dataset *A*. It will estimate the conditional probability for each class, that is, the fraction of points in *A* with that given class label. *I(x)* which is the indicator function evaluates to 1 when the argument *x* is true and *0* otherwise.

In heartwave based biometric classification, the KNN is usually deployed until classification as reported in [21, 22, 67, 68, 80-86]. In a recent work [80], KNN with k=3 as classification method is performed in conjunction with Multiresolution Analysis. Dimension reduction is subsequently performed to determine a subset of features which are highly correlation prior to classification. The method achieved an accuracy of 94.4%. It is to note that the utilized database contains healthy individual whose heartwave signals are acquired under resting condition.

In the reported work of [81], after the segmentation and extraction of heartwave from single lead electrodes, template match methodology is implemented as a prescreening to determine the similarity between the test sample and the templates of 5. The correlated coefficient provides a quantitative measure on the degree of similarity and group the test into one of the template group. Thereafter, Linear Discriminant Analysis distance classification via Euclidean metric system is used to performed classification within the group. To cater for the varying significant of the feature, a weight vector is imposed into the following expression.

$$d(x^p, x^q) = \sqrt{\sum_{i \in R} w_i \left(\mathrm{x}_i^p - \mathrm{x}_i^q\right)^2} \qquad (2.12)$$

where $x^p \; and \; x^q$ are the feature vectors. The smaller the expression of (2.12), the closer the distance between the vectors and hence the distance between the two classes can be computed. As intended by the author, the work is performed on resting heartrate with a total of 168 healthy subjects. It achieved a total of 98% accuracy. It is to note that the work does not cater for heartrate variation which will affect the segmentation and under fitting distance which will result in higher mis-identification.

2.4.2 Neural Network

[66, 83, 87-92] have reported on the use of variants of Neural Network in the classification of the individual based on heartwave signal. Neural Network is essentially a multi layers of nodes consisting of the input layer, hidden layer and output layer. As it is a feed forward neural network (NN), information is only allow to travel from input to output. The input layer represents the raw information that is feed into the network. In the hidden layer, the activity of each unit or node is determined by the activities of the input units and the weights between the input and the hidden units. The output layer is dependent on the activity of the hidden layer and the weights between the hidden and the output. To enable a learned NN, backpropagation is used to optimize the weight values that will lead to minimal total error of the network over the set of training samples. Backpropagation is an iterative process between forward and backward pass. Figure 2.2 illustrates a simple NN.

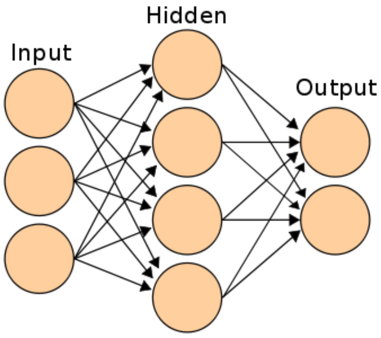


Figure 2.2: Architecture of a generic Neural Network

In [66], heartwave signal is segmented in time domain that involves filtering with various peak detections and autocorrelation. To enable classification, the processed signal from autocorrelation is input to the Artificial Neural Network (ANN) to create a supervised system for individual classification. The ANN is created by initializing settings such as number of Hidden Nodes (set according to the number of users to be registered in the current ANN), number of Output Nodes, Training Ratio (70%), Validation Ratio (15%), Testing Ratio (15%) and the Neural Network name. During the training period, a specific Binary pattern is assigned to each individual, when the autocorrelated input of all user is given to the ANN, it trains itself by varying the weights according to the Back Propagation Algorithm until the assigned binary pattern to each individual is obtained at the output. During the testing period the input of a single user is given to the ANN with the already established weights from the training period, the ANN gives a binary output which is further compared with all the pre assigned binary patterns of the users stored in the database. The user with a match of 80% or above is identified. The work was tested on 4 individuals and achieved an accuracy of approximately 80%.

In another recent work [93], a Convolution Neural Network (CNN) is used to perform user classification. This work used Wavelet Domain Multiresolution Convolution Neural Network (MCNN) methodology to perform user identify without performing heartwave segmentation and extraction. Briefly, it uses wavelet transformation to acquire temporal and spatial information for a fixed length of heartwave signal. For each of the wavelet component, autocorrelation is used to remove blind segmentation-induced phase difference. Thereafter, the data is input into CNN for extraction of feature and supervised learning. Datasets comprising of healthy individual and abnormal signal are used in the testing and the work achieved an average of 93%. To achieve the performance, 50% of the data are used for training the CNN and more importantly, the signals are acquired under resting heartrate condition. Hence, the impact of heartwave morphological changes is not evaluated.

2.4.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is an effective approach that aims to seek directions that are efficient for discrimination as reported in the work [67, 94-98]. Assuming a set of d-dimensional samples that contains multiple subsets of data each labelled as ω_1 and ω_2. The objective of LDA is to project data from d-dimensional space on a line with an expectation to have optimized orientation of the line to achieve minimal error. The mapping and the projection of the samples can be achieved with a linear combination of $y = w^T x$ with the subset of data divided into Y_1 and Y_2. Hence, this lead to the importance of determining **w**. The determination of **w** can be achieved by considering and maximizing the ratio of between class scatter to within-class scatter. Given that the means of the class can be expressed as follows:

$$m_1 = \frac{1}{N_1} \sum_{x_i \in R_1} x_i,\ m_2 = \frac{1}{N_2} \sum_{x_i \in R_2} x_i \tag{2.13}$$

And the project class means can be expressed as follows:

$$\tilde{m}_1 = \frac{1}{N_1} \sum_{x_i \in R_1} w^T x_i = w^T m_1, \qquad \tilde{m}_2 = \frac{1}{N_2} \sum_{x_i \in R_2} w^T x_i = w^T m_2 \tag{2.14}$$

With the difference of the class to be expressed as

$$\tilde{m}_2 - \tilde{m}_1 = w^T (m_2 - m_1) \tag{2.15}$$

The variance of the projected data shall be:

$$\tilde{s}_1^2 = \sum_{y_i : x_i \in R_1} (y_i - \tilde{m}_1)^2 = \sum_{x_i \in R_1} (w^T x_i - w^T m_1)^2 = w^T \left(\sum_{x_i \in R_1} (x_i - m_1)(x_i - m_1)^T \right) w = w^T S_1 w$$

$$\tilde{s}_2^2 = \sum_{y_i : x_i \in R_2} (y_i - \tilde{m}_2)^2 = \sum_{x_i \in R_2} (w^T x_i - w^T m_2)^2 = w^T \left(\sum_{x_i \in R_2} (x_i - m_2)(x_i - m_2)^T \right) w = w^T S_2 w$$

(2.16)

And the ratio of the between class and within class is expressed as

$$r(w) = \frac{(\tilde{m}_2 - \tilde{m}_1)^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{w^T S_B w}{w^T S_w w} \tag{2.17}$$

where
$$S_w = S_1 + S_2$$
$$S_B = (m_2 - m_1)(m_2 - m_1)^T$$

The maximization of the ratio can be achieved

$$w = S_w^{-1}(m_2 - m_1) \tag{2.18}$$

Finally, to achieve classification, based on the condition that if $y(x) > 0$, *x* belong to Y_2 else it belong to Y_1.

$$y(x) = w^T x - \frac{1}{2}(\tilde{m}_1 + \tilde{m}_2) = w^T x - \frac{1}{2} w^T (m_1 + m_2) = S_w^{-1}(m_2 - m_1)(x - \frac{1}{2}(m_1 + m_2))$$

(2.19)

In the work of [94], a fiducial based feature extraction is performed in the time domain through the detection of prominent R-Peak. The subsequent characteristic features are detected about R-Peak through locating the minimum radius of curvature. After the completion of the feature extraction process, LDA is used to perform classification and achieved an average accuracy of 98%. In the work, the heartwave signal comprises of individual under stressful mental states such as reading, mathematical manipulation and driving in virtual reality. The variations of the heartrate under mental stresses are not disclosed and it is extremely unlikely that the heartrate will be as extreme as under physical duress.

2.4.4 Support Vector Machine

The objective of the Support Vector Machine is to construct a hyperplane in a high dimensional space which can be used for classification. This hyperplane is maximized to achieve the largest possible distance between the nearest training data of any class. See Figure 2.3 for illustration.

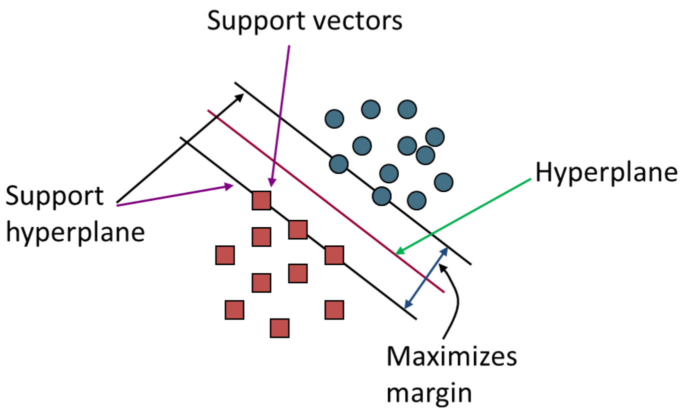


Figure 2.3: Illustration of Support Vector Machine

In linear Support Vector Machine, hyperplane and support hyperplanes can be respectively expressed as follows:

$$\langle w.x\rangle + b = 0 \qquad (2.20)$$

$$\langle w.x\rangle + b = +1 \qquad (2.21)$$

$$\langle w.x\rangle + b = -1 \qquad (2.22)$$

where x is the variable. To determine the optimal separating hyperplane, the distance, d must be as large as possible and it can be computed as follows:

$$d = \frac{1}{\|w\|} \tag{2.23}$$

with the margin as *2d*. Finding the maximum for expression (2.23) is equivalent to finding the minimum of $\|w\|^2/_2$. The margin of the hyperplane is divided by the line from expression (2.20). Hence the class of the point, denoted by *y* is and governed by the following:

$$y_i[\langle w.x\rangle + b] \geq 1\,, i = 1,2,\dots,l \tag{2.24}$$

Where *l* is the number of samples. To solve the optimal value of *w* and *b*, a cost function of can be defined and solved by a Lagrange function. The cost function is defined as

$$\min \emptyset(w) = \frac{\|w\|^2}{2} \tag{2.23}$$

And the Larange function to solve $L(w, b, \alpha)$ is

$$L(w, b, \alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{l} \alpha_i(y_i[\langle w.x\rangle + b] - 1) \tag{2.24}$$

And α is the saddle point for solving the extreme value of the Lagrange function. Support Vector Machine (SVM) as the classifier has been reported in [17, 21, 43, 47, 54, 61, 78, 82, 99-109].

In [54], chaotic theory is implemented in the extraction of the heartwave feature. In this reported work, chaotic theory is adopted in the extraction of 4 chaotic indices which are ECG features, Lyapunov exponent spectrum ($\lambda_1, \lambda_2, \lambda_3, \lambda_4$), correlation dimensions (D_2) and Root Mean Square (RMS)(V_{rms}). Although SVM has been effective for linearly separable data, the reported work does not contain linearly separable data as the data comprises of ECG features, Lyapunov exponent spectrum, correlation dimensions and Root Mean Square. In solving non separable data, the author proposed to re-project the data into higher dimensions before linear separation is adopted. To solve the non-linear problem, non-linear kernel SVM classifier is proposed. The reported work proposed the classification function as shown

$$\max w(\alpha) = max\left[\sum_{k=1}^{l} \alpha_k - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i\, \alpha_j\, y_i\, y_j K\langle x_i \cdot x_j\rangle\right] \quad (2.25)$$

In (2.25), w is maximum distance, l is the number of data points in the training data, y as the output of the data, x as the feature vector in the data set and α as the Lagrangian constant. Lastly, the kernel $K\langle x_i \cdot x_j\rangle$ is the inner product computation between training data which contains non-linear separable data. The author evaluated 3 different kernels namely Polynomial, Radial Basis Function (RBF) and sigmoid based Multi-layer Perceptron. Result shown that the polynomial kernel achieved the highest accuracy. The polynomial kernel is represented by (2.26) with *d* as the degree of polynomial:

$$K\langle x_i \cdot x_j\rangle = \left(\langle x, x^{'}\rangle + 1\right)^{d} \quad (2.26)$$

This work is one of the very few to perform classification for user under exercise. 26 individual signals are acquired under bicycle exercise for a period of 5 minutes and the intensity of the exercise it capped at 35% of the *Target Heart Rate (THR) where it is defined as (Peak HR – Resting HR) x intensity% + Resting HR. Peak HR is defined as (220-age)*. This work achieved a classification accuracy of approximately 81%. The work is commendable but it is unfortunate that the work is not conducted under the full spectrum of heartwave morphological changes. Under extreme physical duress, the intensity of the heartrate can double and triple from resting heartrate.

2.4.5 Generative Model Classifier

A generative classifier is a methodology that relies on Bayes' Theorem. A generative model classifier tries to learn the model that generates the data behind the scenes by estimating the assumptions and distributions of the model. It then uses this to predict unseen data, because it assumes the model that has learned, captures the real model. Generative model is vastly different from discriminative model as discriminative model tries to model by just depending on the observed data. It makes fewer assumptions on the distributions but depends heavily on the quality of the data. There are different variants of generative model such as Gaussian Mixture Model, Hidden Markov Model, Restrictive Boltzman Machine and Probabilistic Context Grammar.

In Bayes' Theorem, it describes the probability of an event based on the prior knowledge of conditions related to the event. Mathematically, Bayes' Theorem is expressed by the following equations.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.27}$$

where A and B are event and the *P(A)* and *P(B)* are the independent events. $P(A|B)$ is the conditional probability of observing event A when B is true, vice versa for $P(B|A)$.

Numerous work has been report on the use of generative model as classifier in classification [16, 22, 46-50, 110]. In a reported work [46], Hidden Markov Model method in conjunction with Hierarchical Dirichlet Auto-regression (HDP-AR-HMM) is implemented to support heartwave feature segmentation and classification. This work uses a combination of discriminative and generative approaches to perform biometric classification. It starts off with the detection of R-Peak which is most prominent. Upon detection of the R-Peak, the segment of the raw data is feed into (HDP-AR-HMM) to map into a sequence of state label by performing joint segmentation and clustering.

After mapping into a string of state label, n-gram language model is used to perform scoring of the loglikelihood. The work is tested on 52 healthy individuals and 238 individuals with cardiac disorder. The performance of the proposed framework achieved an EER of average 25%. Although large dataset has been used, there is no mentioned of data containing heartwave signal under accelerated heartrate.

2.5 Summary

In the elaboration of the current state-of-the-art in the domains of heartwave segmentation and classification, none of the reported work has investigate the performance of their proposed architecture on intense heartwave morphological variations due to physical duress. The current standalone heartwave segmentation methods is not applicable to detect heartwave features under elevated heartrate. However, method like Discrete Wavelet Transformation can be enhanced to aid in the feature extraction. This enhancement will be elaborated in Chapter 3. With regards to the classification techniques, it is essential that the classification techniques cater for variation in discriminating features to enable classification. To recap, the discriminating features under resting condition may be significantly different under elevated heartrate conditions. Classifiers such as Generative Model and Neural Network have the means to cater for variations in discriminating features which will be presented in Chapter 4 and Chapter 5.

CHAPTER 3

HEARTWAVE SEGMENTAION AND FEATURE EXTRACTION

3.1 Introduction

This chapter presents the heartwave segmentation and feature extraction methodology which will allow for accurate and reliable extraction of heartwave characteristic features to support heartwave based biometric classification. The chapter commences with the theoretical understanding of Discrete Wavelet Transformation (DWT), followed by the implementation of DWT integrated with heartrate dependent parameters: PR-Interval and QT-Interval to perform specific waves' components extraction and delineation. The chapter concludes with results of feature extraction and extraction of features from anomaly heartwave signal.

3.1.1 Motivation

- To develop a basis of an extraction method using Discrete Waveform Transformation to partition heartwave signal into various resolution correlated to the frequencies of the heartrate characteristic components.
- To integrate heartrate dependent parameters of PR Interval and QT Interval to extract characteristic features to support feature extraction under extreme heartwave morphological variation.

3.2 Proposed Method of Discrete Wavelet Transformation with Heartrate Dependent Parameters

To achieve a reliable segmentation and extraction of heartwave characteristic features, the methodology proposed the use of Discrete Wavelet Transformation (DWT) together with heartrate dependent parameters of QT-Interval and PR Interval. In variable heartrate, the amplitude and the temporal locations of the characteristic features displaced is proportion to the heartrate. To address the challenges of locating the features, a fixed window length for detection of features reported by numerous work is not applicable as it leads to false detection in highly elevated heartrate.

Briefly, the proposed methodology commences with using DWT to decompose the signal into variable levels. The intent is to isolate heartrate signal into different bands of frequency. DWT technically comprises of high pass and low pass filter. The decomposition of signal into various levels allows different components of the heartwave to be easily detected using peak and valley detection function. Upon decomposition, R-Peak is the first characteristic point from QRS-Complex to be identified. This is followed by detection of Q-Valley and S-Valley of QRS-Complex. To address the heartrate variation, the detected R-Peaks are used to determine the R-R interval in each sequence. Converting R-R interval to heartrate, QT-Interval and PR-Interval can be computed. PR-Interval is imposed from the left region of R-Peak to detect the components of P-Wave components. QT interval is imposed from Q-Valley into the right region of R-Peak to detect the components of T-Wave components. See Figure 3.1 for illustration of the methodology on heartwave characteristic features detection.

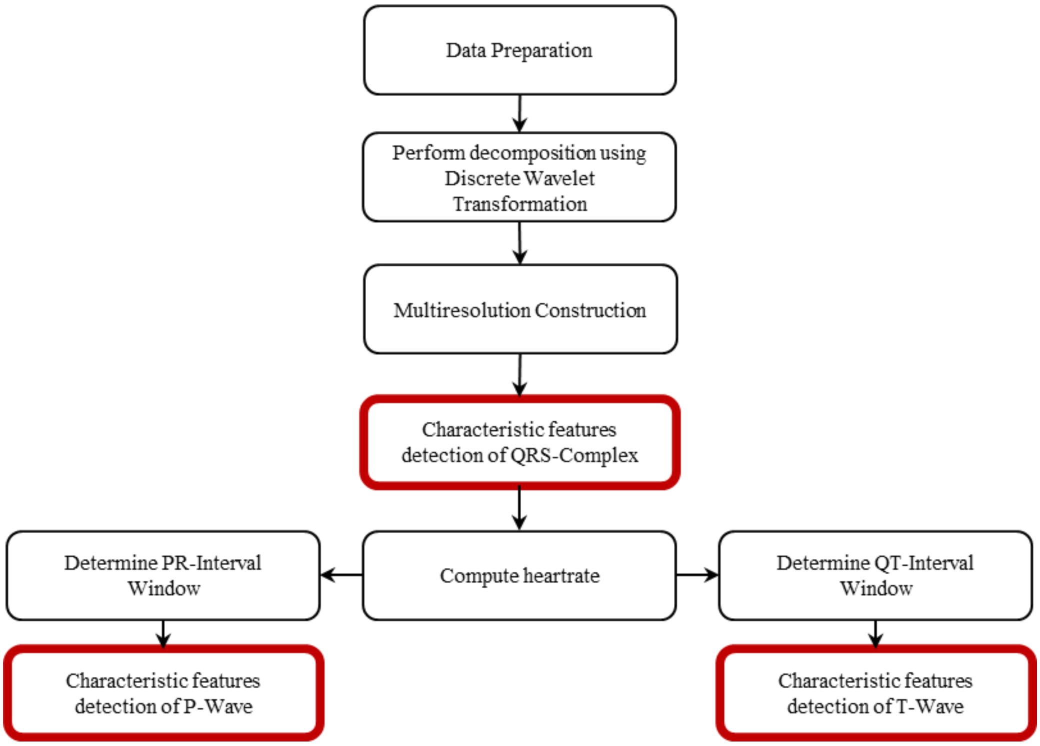


Figure 3.1: Proposed methodology for heartwave characteristic features detection

3.2.1 Data Preparation

To prepare for characteristic features detection, the dataset from individual needs to be prepared. For each heartwave, a total of 11 unique features are to be extracted. To maintain consistency in the detection, only Lead I and Lead II signals will be used. In the data preparation, the data from each individual is segmented into sequences of 10 seconds duration. Each of the sequences can contain from 8 to as much as 30 individual segmented sequential heartwaves which is proportional to the heartrate acquired at that instance. See Figure 3.2 for illustration of data preparation. Each of the sequence is input to the proposed methodology to extract the features. The extracted features are compiled into a dataset for each individual. In total, there are 27 records of heartwaves signals acquired under treadmill testing. The duration of the treadmill test ranges from 13 minutes to as long as 67 minutes. Importantly, the minimum heartrate which is the resting heartrate varies from 50 beats per minute (bpm) to as high as 90 bpm and the maximum heartrate ranges from 82 bpm to as high as 190 bpm. The extremities that an individual can stretch from their resting heartrates to maximum heartrates ranges from 150% to more than 300%.

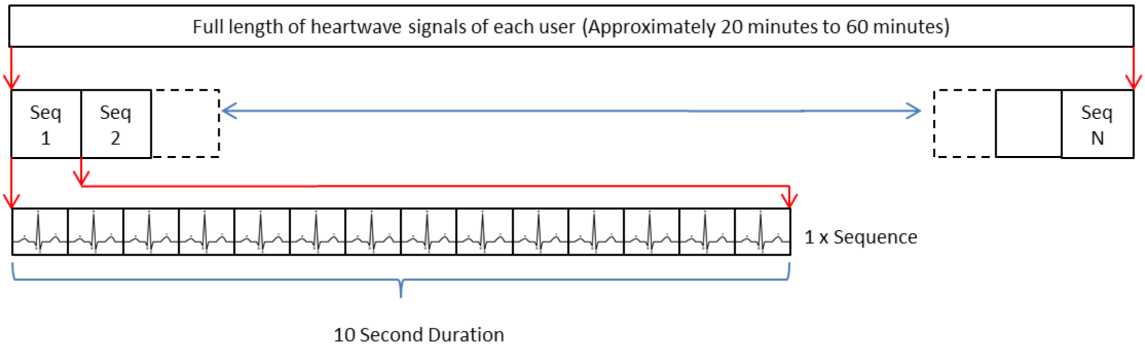


Figure 3.2: Schematics of data preparation for feature extraction

3.2.2 Introduction of DWT

DWT has been widely used to separate signals into different frequency bands where critical information at different scales can be easily qualified and quantified. Advantages of DWT methodology is the ability to determine the time location of the desired frequency. Wavelet analysis consists of signal decomposition (wavelet transform) and reconstruction (inverse wavelet transform) phases. Alfred Haar is credited with the first use of wavelet, although the term wave was not coined until Morlet used it in his signal processing work of 1983. The wave shape used by Haar is known as the Haar wavelet. It is the simplest wavelet possible and resembles a step function. The Haar wavelet is discontinuous and therefore consequently yields poor frequency localization. Haar wavelets are helpful for developing a basic understanding of wavelet analysis but not often use in practice. Ingrid Daubechies constructed wavelet bases with compact support which implies that the wavelets are non-zero on an interval of finite length. Compact supported wavelet families accomplish signal decomposition and reconstruction using only Finite Impulse Responses (FIR) filters. Stephane Mallet proposed the concept of Multi-Resolution Analysis (MRA). MRA is based on the concept that objects can be examined using varying levels of resolution. An example is the zooming concept in viewing graphics. When the view is a distance from a picture, a low level of details can be used to represent the object. As the user zooms into the picture, the resolution of the picture increases with higher details in sight. This zoom-in zoom-out property of MRA serves as one of the basic properties for wavelet analysis. Importantly, DWT has the characteristic of orthonormal wavelet. Orthonormal wavelets have specific properties that provide a means of efficient decomposition of a signal. Orthonormal wavelet function defines a specific set of filters for efficient signal decomposition and reconstruction.

3.2.3 Theory of Discrete Wavelet

In DWT, wavelet methodology analyze the input signal in section by translation of a analysis function with a wavelet function, ψ. The wavelet function is scaled (expanded or dilated) in addition to be translated in time. The ψ is often called a mother wavelet because it “gives birth” to a family of wavelets through the dilations and translation. Given a mother wavelet ψ, an orthonormal basis of $\{\psi_{a,b}(t)\}$ in the Hilbert space of $L^2(R)$, the square of square integrable function is defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi * (\frac{t-b}{a}) \tag{3.1}$$

Where a represents the scale and b represents the translation parameters. The scale parameter, a, indicates the level of analysis. Small values of a provides a local, fine gran or high frequency analysis while large values correspond to large scale, coarse grain or low-frequency analysis. Changing the b parameter moves the time localization center of each wavelet; each $\psi_{a,b}(t)$ is localized around $x=b$.

In DWT, the scale factor between levels increases by two. Thus, scaling is also known as dilation. Widely used a and b parameters setting create an orthonormal bases as $a=2^j$ and $b=2^j k$ $(j,\ k\ \in z)$ where j refers to the level and k as the translation index. The wavelet family from the substitution of a and b into expression (3.1), becomes

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right) \tag{3.2}$$

$\psi_{j,k}$ is an orthonormal basis because $\int\left(\psi_{j,\mathrm{k}} \cdot \psi_{l,\mathrm{m}}\right) = 0$, $((j \neq l)(k \neq m))$ and the norm of unity for L^2 , $\int\left(2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right)\right)^2 = 1.$ The wavelet transform calculates the wavelet coefficients by taking the inner product of an input signal, *f(X),* with a wavelet function $\psi_{j,k}(t)$ which leads to the Discrete Wavelet Transformation (DWT) of the following form:

$$D_{j,k} = \langle f, \psi_{j,k}\rangle = 2^{-j/2}\int_{-\infty}^{+\infty} f(x)\ \psi(2^{-j}x - k)dx \tag{3.3}$$

where $D_{j,k}$, are the wavelet coefficients. The wavelet coefficients are measures of the goodness of fit between signal and the wavelet. Large coefficients indicates a good fit.

3.2.4 Multi-Resolution Analysis

Calculating wavelet expansion from the expression (3.3) directly is computationally intensive. Mallet introduced the technique called Quadrature Mirror Filtering which shows that any discrete wavelet transformation can be calculated rapidly using cascade-like algorithm. This reduces the number of operations necessary for the translation and transformation. Technically, a Multi-Resolution Analysis projects a function on a set of closed subspaces.

$$V_{-2} \subset V_{-1} \subset V_0 \subset V_{+1} \subset V_{+2} \subset V_{+3} \tag{3.4}$$

With the nested subspaces, V_j can be expressed as the direct sum of the approximated subspace V_{j-1} with the detail subspace, W_{j-1}:

$$V_j = V_{j-1} \oplus W_{j-1} \tag{3.5}$$

The function $f_N \in V_N$ can be split into different components represented by subspaces $V_{N-i}\,(i = 1, 2, \ldots)$ together with the orthogonal complements through MRA:

$$f_N = f_{N-1} + g_{N-1} = \sum_i^M g_{N-M} + f_{N-M} \tag{3.6}$$

At each level, the scaling function changes. Based on the example of zooming picture, subspaces V_j can represent the zoom-in effects at different level and with each level, different details are manifested. Hence the expression

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1},\ j \in Z \tag{3.7}$$

The pairs of functions f_{N-1} and g_{N-1} can be obtained by applying a pair of Quadrature Mirror Filters (QMR) to the original function f_N. QMR is a filter most commonly used to implement a filter bank that splits an input signal into two bands. The resulting high-pass and low-pass signals reduce the signal by a factor of 2, giving a critically sampled two channel representations of the original signal. For DWT, the QMR are a pair of sequences, $\{h(k)\}$ as the low pass filter and $\{g(k)\}$ as the high pass filter. The low-pass filters averages the data and the high-pass filter contains more information. The relationship between the two filters are expressed as

$$g(n) = (-1)^n h(1-n) \tag{3.8}$$

At each of the space of V_0, the signal is transform by the scaling function, ϕ. This implies that for the subspace, the functions can be combined linearly for the next level of subspace V_1. Hence the relationship for the scaling functions between adjacent subspaces, V_j and V_{j+1} is known as the scaling/dilation equation which defines the filter coefficients.

$$\phi_{j,k}(x) = \sum_n h_\phi(n)\phi_{j+1,n}(x)$$

Substituting $\phi_{j,k}(x) = 2^{\frac{j}{2}}\,\phi\left(2^j x - k\right)$ into the equation above

$$\phi_{j,k}(x) = \sum_n h_\phi(n) 2^{-\frac{j+1}{2}} \phi_{j+1,n}(2^{j+1}x - n)$$

$$\phi(x) = \sum_n h_\phi(n)\sqrt{2}\phi(2x - n) \tag{3.9}$$

Similarly the wavelet may be refined with double resolution as shown below

$$\psi(x) = \sum_n g(n)\sqrt{2}\phi(2x - n) \qquad (3.10)$$

where the coefficients g_n are useful in wavelet decomposition as the highpass filter.

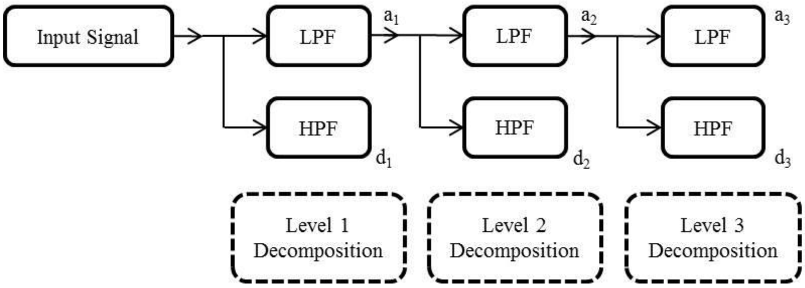


Figure 3.3: Illustration of a Multi-Resolution Analysis

From (3.9), the equation transform a vector with *n* elements into two vectors with length of *n/2* each. One of which contains the data smoothed by the low-pass filter with the other containing the details. See Figure 3.3 for illustration. Each wavelet is characterized by a finite sets of filter coefficients derived from the scaling equation that relates scaling functions of different subspaces, V_i, to each other. For Daubercies 4th order wavelet, the coefficients for Scaling and Daubechies DB4 wavelet are as follows

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \qquad (3.11)$$

$$g_0 = \frac{3-\sqrt{3}}{4\sqrt{2}}, g_1 = \frac{-(3-\sqrt{3})}{4\sqrt{2}}, g_2 = \frac{3+\sqrt{3}}{4\sqrt{2}}, g_3 = \frac{-(1+\sqrt{3})}{4\sqrt{2}} \qquad (3.12)$$

Notice that (3.11) and (3.12) are bounded by the relationship as defined in (3.8). For convenience, low-pass and high-pass filters are denoted by H and G with the sequence represented by $a=\{a_n\}$

$$(Ha)_k = \sum_n h(n-k)a_n \tag{3.13}$$

$$(Ga)_k = \sum_n g(n-k)a_n \tag{3.14}$$

To appreciate the wavelet transformation, let the data be represented by $c^{(n)}$, which is a vector with 2^n elements. Letting $c^{(n-1)} = Hc^{(n)}$ and $d^{(n-1)} = Gc^{(n)}$. Through the application of low-pass filtering, it will double $c^{(n-2)} = H^2c^{(n)}$ and $d^{(n-2)} = HGc^{(n)}$. Using MRA, the DWT output of a sequence $y = c^{(n)}$ with a length of 2^n is equivalent to another sequence of the same length represented by

$$w = \left[d^{(n-1)}, d^{(n-2)}, \dots, d^{(1)}, d^{(0)}, c^{(0)}\right] = [Gy, GHy, GH^2y, \dots, GH^{n-1}y, GH^ny, H^ny] \tag{3.15}$$

(3.15) is the wavelet transform consists of all layers of details, from fine to coarse, stacked next to each other. Figure 3.4 shows the graphical representation of the multiresolution analysis.

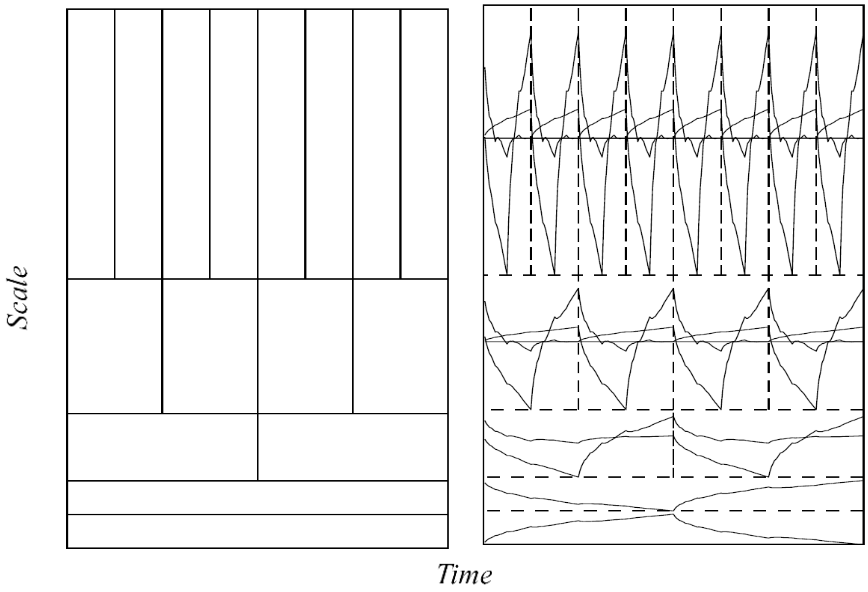


Figure 3.4: Illustration of a Multi-Resolution Analysis on Daubechies DB 4 Wavelet

3.2.5 Heartwave Feature Extraction

The extraction of heartwave features uses DWT with the fourth order mother wavelet of Daubechies. None has reported on the use of DWT for extraction of heartwave features where heartrate of individual is under physiological duress. Daubechies DB 4 has been chosen as it has a close resemblance to the heartwave signal. Figure 3.5 shows the architecture of the heartwave features detection methodology. Broadly, the extraction is partitioned into three stages. The first stage centered on the feature extraction of QRS-Wave complex. Second stage focuses on the extraction of features of P-Wave. The last stage focuses on the extraction of features of T-Wave.

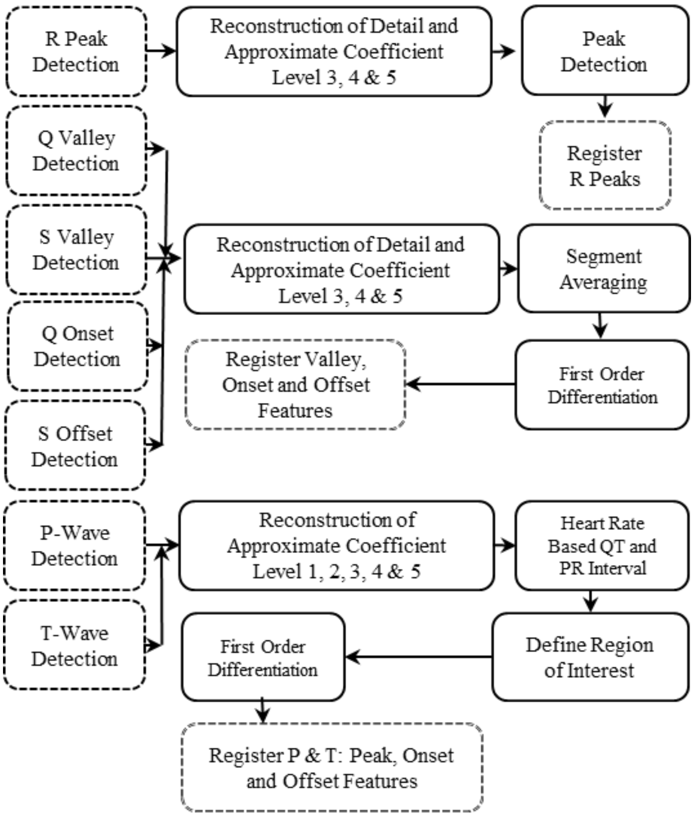


Figure 3.5: Process architecture for feature extraction

3.2.5.1 QRS-Wave Complex Extraction

3.2.5.1.1 R-Peak Detection:

In the detection of the heartwave characteristic features, R-Peak detection is the most essential as it establishes the reference for detection of other characteristic features. The R-Peak detection starts with the removal of baseline signal of 50Hz which is contributed by the mains AC electrical signal. Thereafter the signal is decomposed using Discrete Waveform Transformation with mother wavelet of Daubechies 4. In the determination of R-Peak, Detail Coefficients from Level 3, 4 and 5 are reconstructed and the frequency spectrum from the reconstructed coefficients exhibited a clustering of frequencies between 15Hz to 25Hz. This reconstruction of signals using multiple levels is in contrast to the work by Mazomenos [111] which uses only one level. This reconstruction of multiple levels allows the extraction of heartwave features to be spanned across the full spectrum of the heartrate from resting to maximum heartrate of an individual. Hence, using only one-level from DWT is significantly inadequate to perform a reliable feature extraction. Squaring of the reconstructed signal is implemented to remove noise in the negative region. Using maximum peak detection algorithm, the peaks in the reconstructed signal are detected. See Figure 3.6 on the proposed methodology for R-Peak detection. As the peaks are not uniform in amplitude, a 40% threshold factor against the maximum peak is proposed.

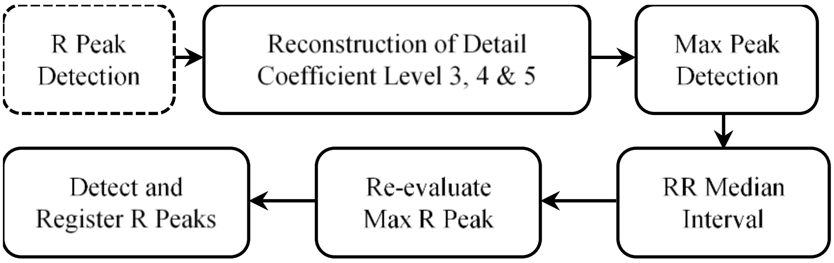


Figure 3.6: Block diagram for R-Peak detection

In typical ECG signal acquisition, due to placement of ECG acquisition sensor and motion artifact, there are uncommon occurrence of spikes between consecutive R peaks. See Figure. 3.7 for details. Although amplitude threshold has been established at 40% of maximum peak, the random spikes can still be detected. Hence, an effective approach is the determination of time period between peaks. Thereafter, a median ranking is performed. As random spike is not a common occurrence, in event of random spike, the false time period between 2 peaks will be ranked either in the front or towards the end. Hence through the use of median ranking of time period, the approximated true time period can be extracted. This approximated true time period can be used to narrow the region of search of true R-peaks with successful omission of random spike.

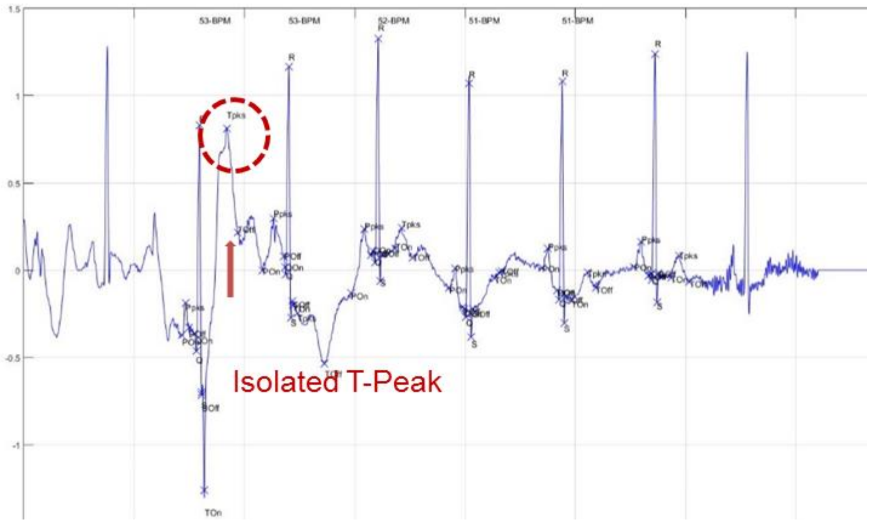


Figure 3.7: Spike signal between two R-Peaks

3.2.5.1.2 Q-Wave Peak and S-Wave Peak Detection:

After the completion of R-Peak detection, the algorithm proceeds to detect Q-Peak from Q-Wave and S-Peak of S-Wave. The detection process uses the reconstructed signal comprises of detail coefficients from Level 2, 3, 4 and 5. The resultant reconstructed signal is a filtered signal void of high frequency components. Q-Peak and S-Peak which are the minimal points about R-Peak use established method available in Matlab to detect the mentioned peaks.

3.2.5.1.3 Q-Wave Onset and S-Wave Offset Detection:

In most published papers reviewed which performed heartwave characteristic features extraction to support heartwave based biometric authentication, the definition of QRS-wave complex has been defined from Q-Wave peak (local minima) to S-Wave peak (local minima) [57, 59, 61, 112] except [56] which uses modulus maximum with Continuous Wavelet Transform for onset and offset detection. This definition is inconsistent from medical definition where QRS-Wave complex is defined from Q-

Wave onset to S-Wave offset [113]. The determination of Q-Wave onset is traced and determined using first order differentiation from Q-Wave peak. See Figure 3.8 for schematic process flow.

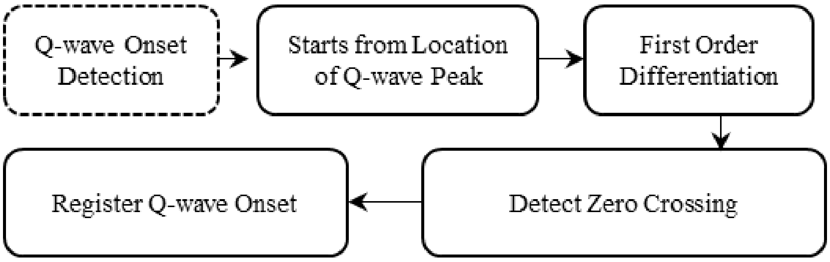


Figure 3.8: Block diagram for Q-Wave onset detection

The first order differentiation approach is the one-dimensional 5-point stencil numerical differentiation [57] that performs first derivation of the point of interest with 4 other adjacent values using the formula as shown in (3.16).

$$f'(x) \approx \frac{-f(x+2h)+8f(x+h)-8f(x-h)+f(x-2h)}{12h} \tag{3.16}$$

where *x* is the point of interest and *h* represents the adjacent values of *x*.

Q-Wave onset is the inflexion point also knows as the stationary point that lies between the falling edge of P-Wave offset and rising edge of the Q-Wave peak. The first order differentiation is adopted within a defined segment of the signal with a period of 100msec. Thereafter a zero-crossing detection is performed. All the values from the first order differentiation process within the defined segment are evaluated. The first value that crosses from positive region to negative region is defined as the first zero crossing which is the inflexion point and the P-Wave onset is registered. See Figure 3.9 for illustration.

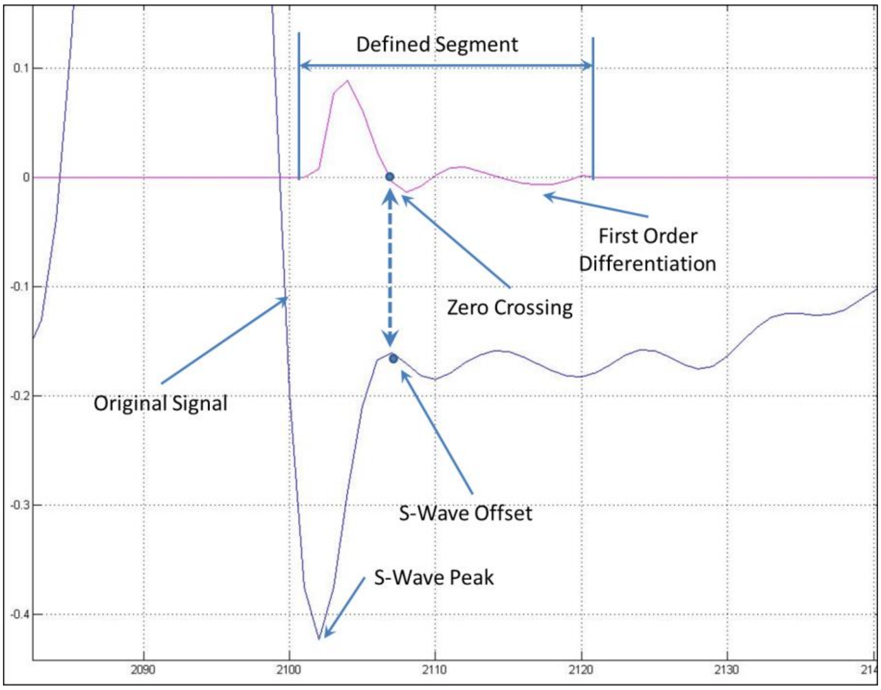


Figure 3.9: Example of an inflexion point on S-Wave Onset

While first order differentiation can be efficient, the method has a drawback. The presence of random noise can degrade the signal-to-noise ratio of first order signal. This can result in many false zero-crossings. To overcome this limitation and to maintain the integrity of the whole signal, a signal smoothing operation is adopted within a defined segment as mentioned in the earlier paragraph. A polynomial based curve fitting of up to 6th degree is used as an approximate fit to the signal segment using the polynomial equation as shown below:

$$y = ax^6 + bx^5 + cx^4 + dx^3 + ex^2 + fx + g \tag{3.17}$$

where *a, b, c, d, e, f, and g* are coefficients for 6th degree polynomial curve fitting. The values of the coefficients are dynamic and dependent on the variability of the signal portion. The coefficients can be determined using Cramer's Rule. Thereafter, a first order derivative using 5-point stencil numerical differentiation is performed followed by zero-crossing detection. See Figure 3.10 for illustration.

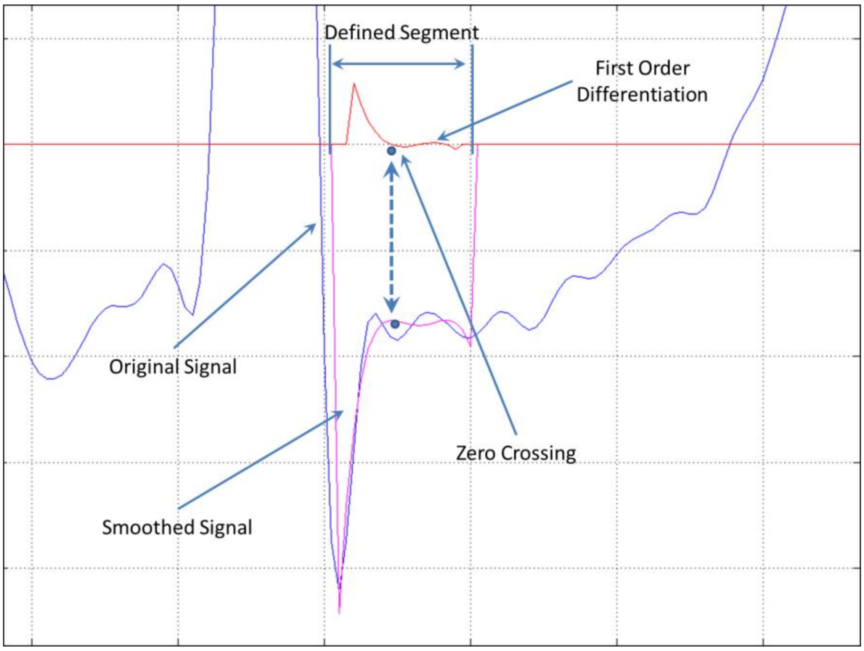


Figure 3.10: Detection of inflexion on smoothened signal

The determination of S-Wave offset is similar to the method described in the detection of Q-Wave onset. S-Wave offset is almost symmetrically located about R peak after S-Wave peak. However, algorithm is enhanced to support heartwave signal of different ECG electrodes from Leads of aVL, V1, V2 and V3. Leads of aVl, V1, V2 and V3 produce high negative amplitude of S-Wave. Using the similar method to detect Q-Wave onset, characteristic feature S-Wave offset can be easily detected. See Figure 3.11 for details.

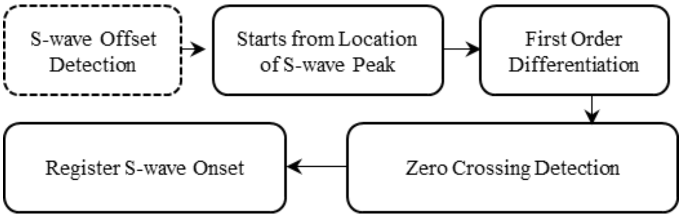
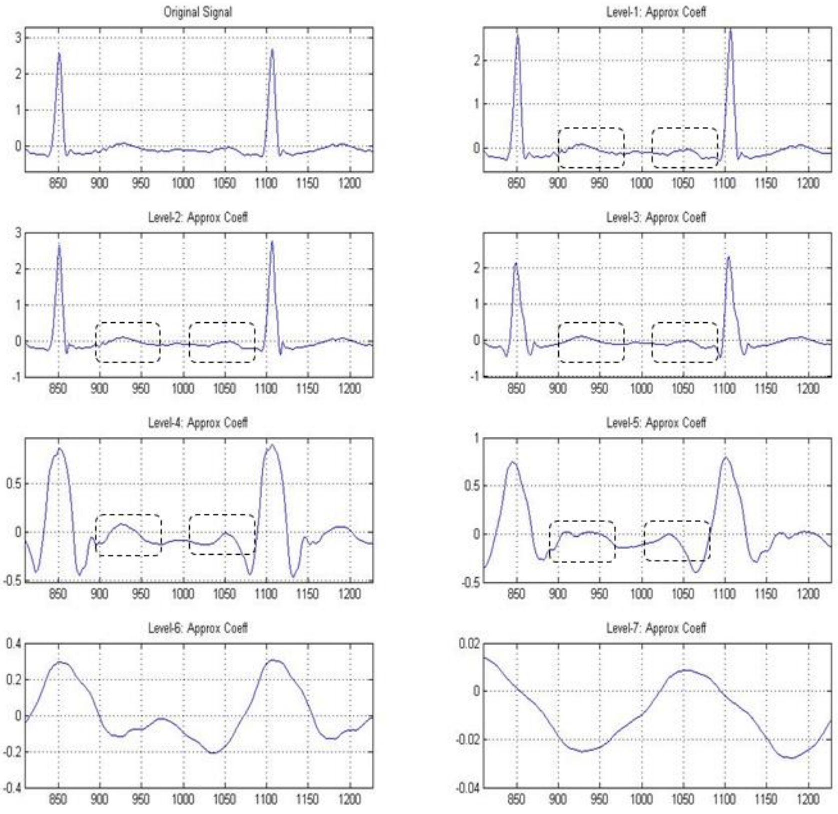


Figure 3.11: Approach for S-Wave offset detection

3.2.5.2 T-Wave Components Extraction

3.2.5.2.1 T-Peak Detection

A simple and novel method for the detection of T-Wave component has been proposed. The method uses approximate signals of Discrete Waveform Transformation. This is in contrast to all the mentioned approaches proposed by others where detail signals have been used. In T-Wave peak detection, the detection leverages on the approximate signals of Discrete Waveform Transformation of Level 1, 2, 3, 4 and 5. See Figure 3.12 for details. As high frequencies components have been filtered, the resulting combination of approximation signals contains relatively clean signal. Frequency spectrum analysis of approximate signals from Level 1 to 5 shows clustering of frequencies between 2Hz to 42Hz.



Original Signal

Level-1: Approx Coeff

Level-2: Approx Coeff

Level-3: Approx Coeff

Level-4: Approx Coeff

Level-5: Approx Coeff

Level-6: Approx Coeff

Level-7: Approx Coeff

Figure 3.12: Distinct profiles of T-Wave and P-Wave from DWT approximate signal of Level 1 to 5

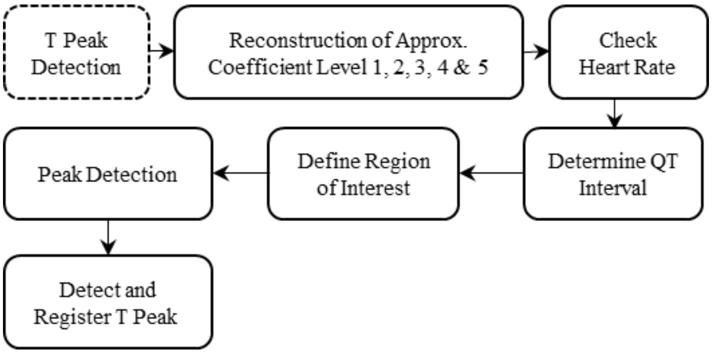


Figure 3.13: Block diagram for T-Wave peak detection

Using the combined approximate signals, an adaptive region of interest is defined to enable the detection of the local maxima. See Figure 3.13 for details. The region of interest is a dynamic window that is dependent on the heart rate of the individual. The heart rate dependent window takes reference from the QT interval which is defined from the onset of Q-Wave and terminates at the offset of T-Wave. QT interval duration is inversely proportional to the heart rate [37, 45]. The region of interest based on QT Interval Duration is derived from the QT Interval Nomogram [114, 115] which is a clinical risk assessment tool that predicts risk of QT prolongation in individual in respond to cardiac related drugs. The region of interest based on QT Interval Duration in accordance to QT Nomogram is determined with the following equation:

$$QT\ Duration\ (msec) = 2.2095\ \times HR_{BPM} + 627.41 \tag{3.18}$$

The equation for QT Duration is valid for heart rate that falls within the range from 64bpm to 154bpm. For heart rate below 64bpm, a constant QT Duration of 484msec is defined. Figure 3.14 shows the consistence of the heartrate dependent QT Interval under variable heartrate.

During the testing of the signals acquired under exercise duress, there are encounters where the heart rate of individual are in excess of 150bpm to as high as

180bpm. For heart rate at rate greater than 150bpm, an occurrence of Supraventricular Tachycardia is observed. See Figure 3.15 for details. The occurrence of Supraventricular Tachycardia in healthy individuals in context of testing is attributed by the physiological stress from ECG treadmill. In situation where heart rate is in excess of 150 beats per minute coupled with the occurrence of Supraventricular Tachycardia, P-Wave delineation is almost impossible in Lead I electrode as T-Wave eclipsed the P-Wave of succeeding heartwave. During this instance, the QT Interval Duration is increased by 20% of maximum QT duration at heart rate of 154 bpm to include the morphology of P-Wave of successive heartwave. Thereafter, a local maximum is performed to detect the T-Wave peak and the detected T-Wave peak can be registered.

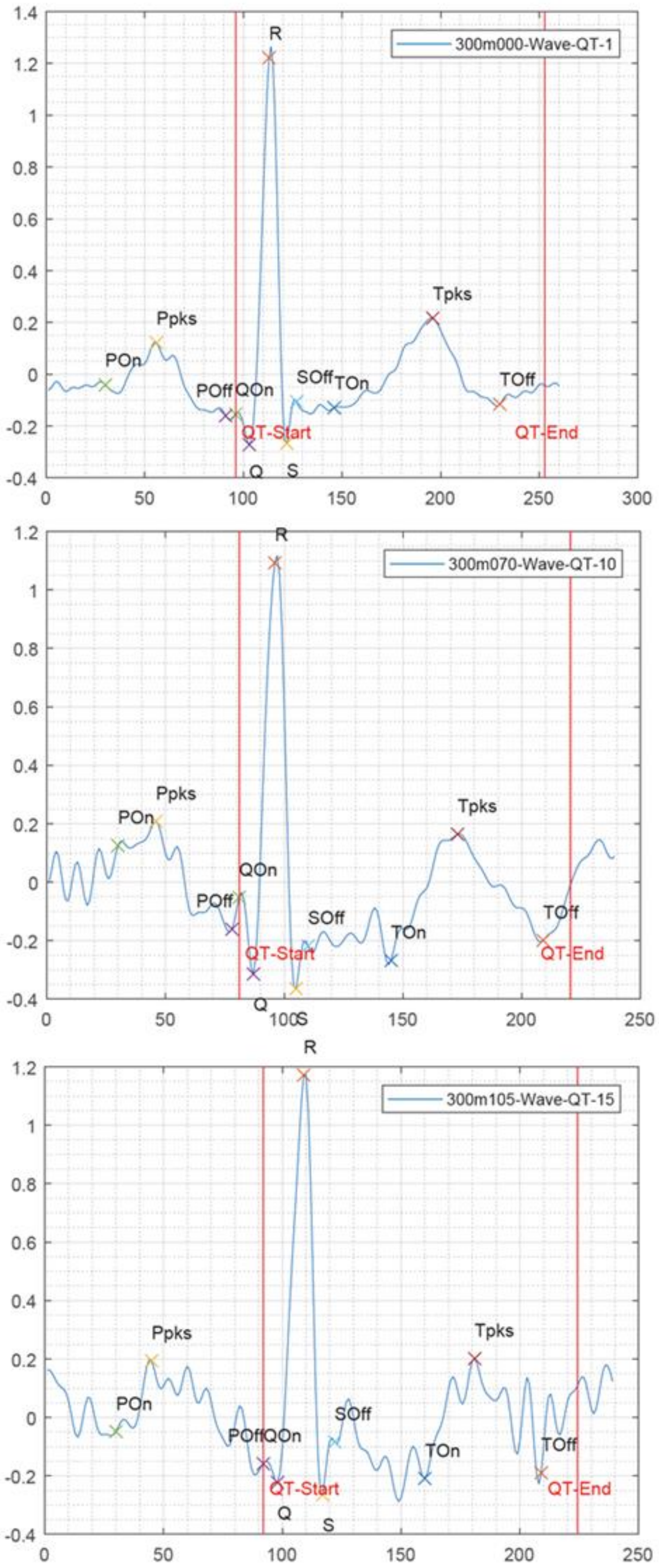


Figure 3.14: QT Interval window at various heartrate

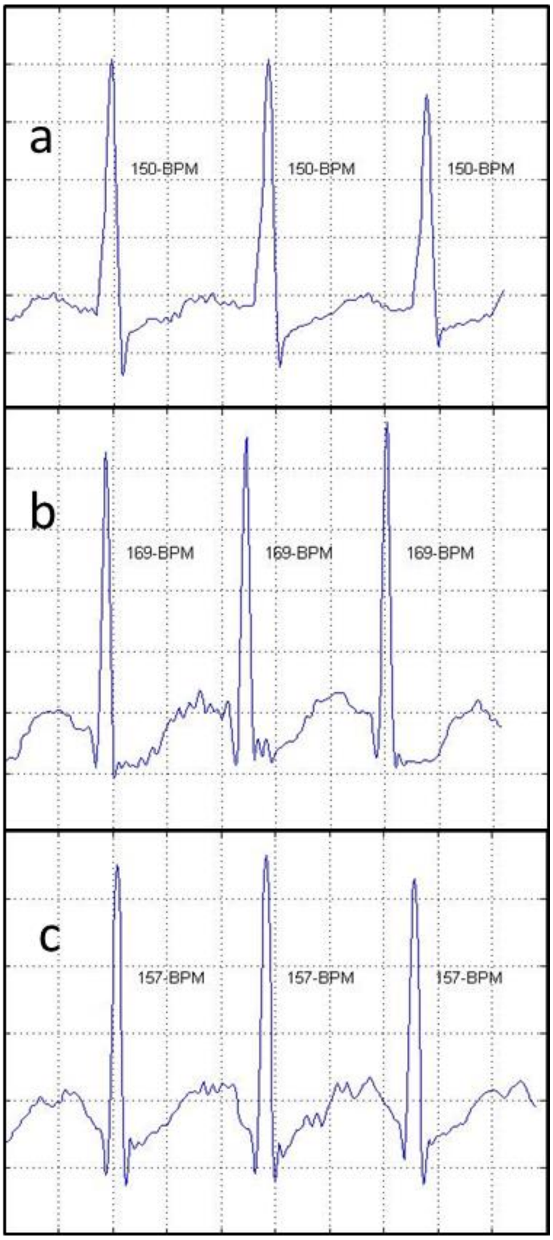


Figure 3.15: Occurrence of Supraventricular Tachycardia of three different individuals at respective heart rate a. 150 bpm, b. 169 bpm and c. 157 bpm. Notice the non-existent of P-Wave (eclipsed by T-Wave)

3.2.5.2.2 T-Wave Onset and Offset Detection:

Upon the successful detection of T-Wave peak, the same reconstructed signal comprises of approximate signals from Level 1, 2, 3, 4 and 5 is used to perform T-Wave offset and onset detection. The reconstructed signal provided a clear outline of the T-Wave that comprises of the local maxima and two local minima located about the local maxima that represent the T-Wave onset and offset respectively.

Local minima detection together with zero crossing detection using first order differentiation is conveniently applied to detect T-Wave offset and onset. See Figure 3.16 for details. For T-Wave onset, the region of interest is established from the presiding characteristic feature S-Wave offset to T-Wave peak. This region of interest ensures and minimizes wrong detection of the local minima. With regards to the T-Wave offset, the region of interest relies on the adaptive boundary window determined from QT Interval Duration as defined in accordance to the QT Nomogram Plot [114].

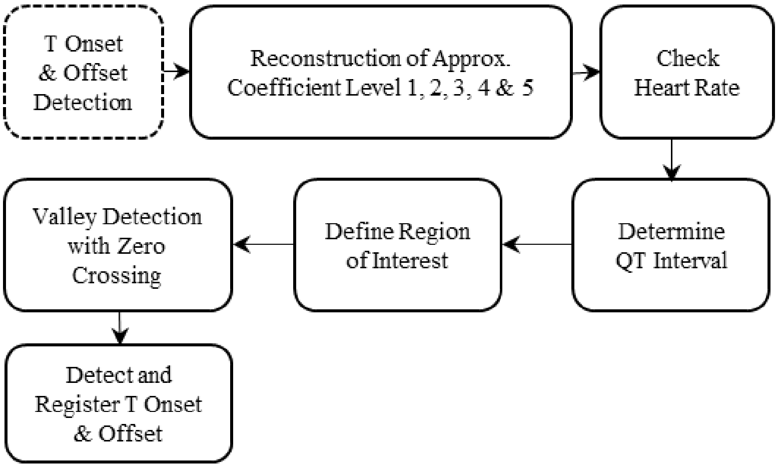


Figure 3.16: Block diagram for T-Wave onset and offset detection

3.2.5.3 P-Wave Components Extraction

3.2.5.3.1 P-Peak Detection:

The detection of P-Wave peak is similar to the detection of T-Wave peak where the reconstructed signal comprises of approximate signals from Level 1, 2, 3, 4 and 5. Similar to T-Wave, P-Wave duration is dynamic and is inversely proportional to heart rate[39]. Hence an adaptive region of interest for detection of P-Wave is adopted. As P-Wave resides on the left portion of the R-peak, the region of interest utilizes the PR

Interval Duration instead of the QT Interval Duration. PR Interval Duration is a dynamic interval that varies according to the rate of the heart rate. PR interval is used medically to diagnose cardiac related issue. According to medical definition, PR Interval is defined from the onset of P-Wave till the onset of Q-Wave. It is affirmative that the termination is at onset of Q-Wave contradicting the implied name of PR Interval. The region of interest is defined using the equation developed by [39, 116, 117]

$$PR\ Interval\ (msec) = -0.351\ \times HR_{BPM} + 176.7 \tag{3.19}$$

The PR Interval is valid for heartrate that falls within the range of 60bpm to 160bpm. Upon the determination of region of interest for detection, local maxima detection is employed to facilitate the detection of the P-Wave peak. The detected P-Wave peak is detected and registered. See Figure 3.17 for details.

In the event of heart rate in excess of 150 coupled with the occurrence of Supraventricular Tachycardia, the equation is invalidated and superseded the calculated PR Interval Duration at maximum heart rate of 160 bpm increases by 20%. As mentioned in T-Wave peak detection, P-Wave in Supraventricular Tachycardia occurrence cannot be delineated in Lead I electrode as it is being eclipsed by preceding T-Wave and hence the assumption of values from preceding T-Wave.

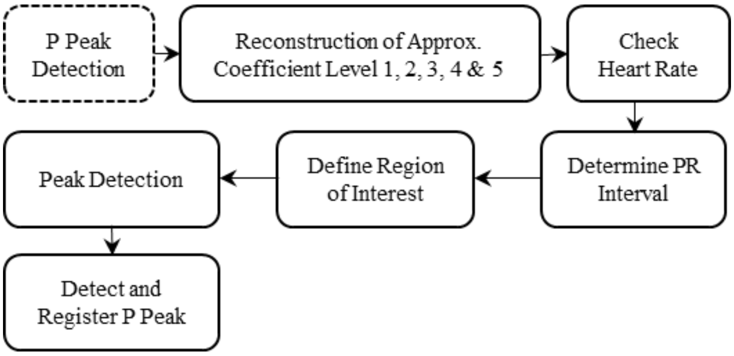


Figure 3.17: Block diagram for P-Wave peak detection

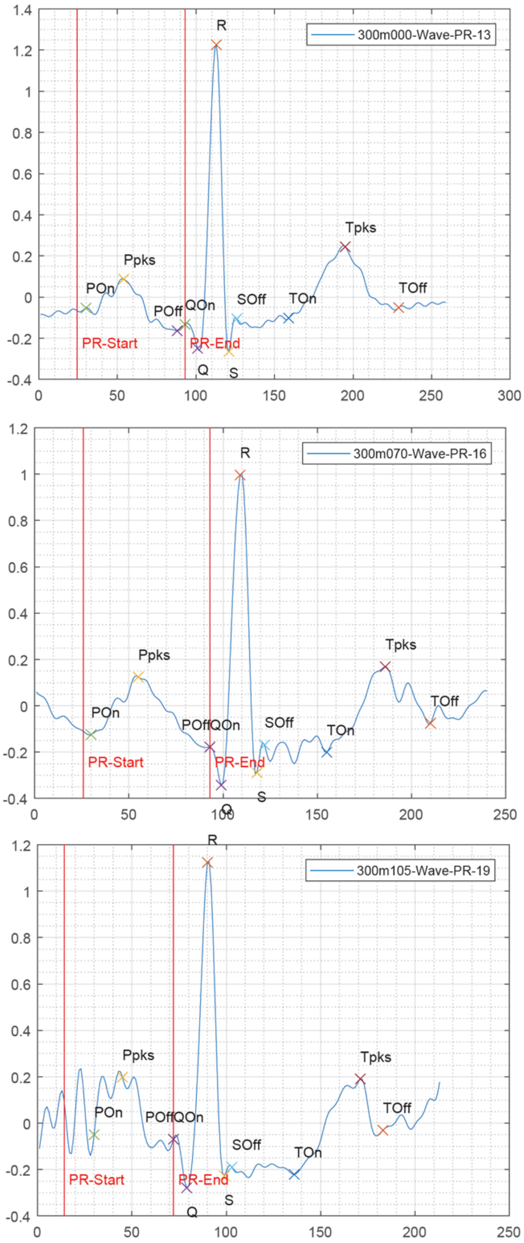


Figure 3.18: PR Interval window at various heartrate

3.2.5.3.2 P-Wave Onset and Offset Detection:

The detection of P-Wave onset and offset is highly similar to the detection of T-Wave onset and offset. The reconstructed signal comprises of the approximate signals of Level 1, 2, 3, 4 and 5. Detection of onset and offset are detected using local minima coupled with the tracing of zero crossing using first order differentiation which detects the deflection to and from P-Wave peak. See Figure 3.19 for details.

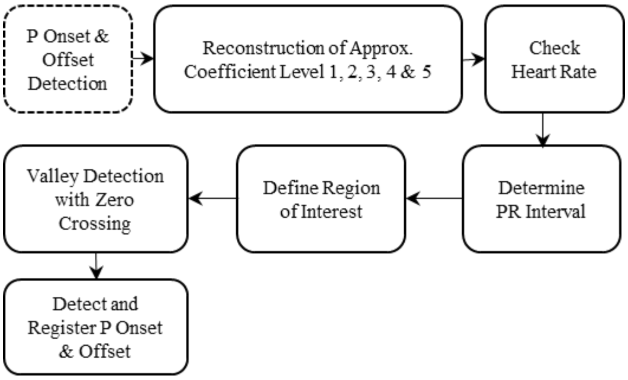
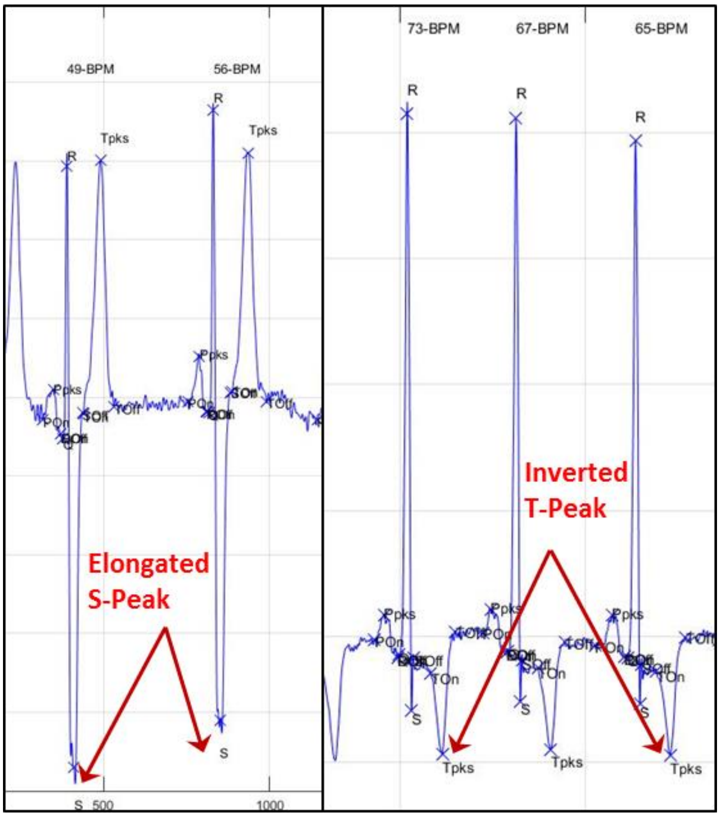


Figure 3.19: Block diagram for P-Wave onset and offset detection

3.2.5.4 Extraction of Signals from Unipolar Chest:

The algorithm is developed to perform extraction of the features from Lead I. Lead I is known as lateral lead where the signal is acquired from both arms. Incidentally, in the course of developing the approach for features extraction, the database contains signals from unipolar chest leads namely V1, V2 and V3 and signal with inverted T-Wave.

Lead V1, V2 and V3 are leads acquired from the chest of an individual. The characteristic feature of the mentioned leads is the significant S-Peak of the QRS-Wave complex. See Figure 3.20a for details. In addition to the ultra-high amplitude spike, the signal exhibits an extended duration from S-Peak to onset of T-Wave. This phenomenon is due to the repolarization of the heart muscle that follows after contraction and depolarization. During repolarization, the cardio muscle elongates to prepare for the subsequent heartbeat and it takes longer than the previous depolarization.



(a) (b)

Figure 3.20: Successful extractions of features of heartwave signals with: (a) elongated S-Peak and (b) inverted T-peak.

Incidentally during the development, detection of inverted T-Wave signal has also been successfully implemented. See Figure 3.20b for details. Except for signal acquired from Augmented Vector Right (aVR) lead, where the signal is acquired from right wrist reference to the ground lead, the T-Waves from all leads are typically in upright position relative to the position and direction of QRS-Complex. Medically, inverted T-Wave has been extensively studied and there is no definite correlation of inverted T-Wave to cardiac related symptoms. Inverted T-Wave has been found in healthy individuals[118]. Although uncommon, the algorithm has been enhanced to support the feature extraction for signal with inverted T-Wave. As the T-Wave is inverted, it is necessary to polarize inverted T-Wave signal into positive region.

3.3 Experimentation and Results

The algorithm has been tested on public databases: MIT-BIH ST Change from Physionet. The database contains readings of 27 individuals with most recordings recorded with a minimum duration of 20 minutes. Of the 27 readings, 5 reading belongs to individual with anomaly heartwave signal. A total of 63,388 heartwaves have been fed to the methodology and the feature characteristics are successfully extracted as shown in Table 3.1.

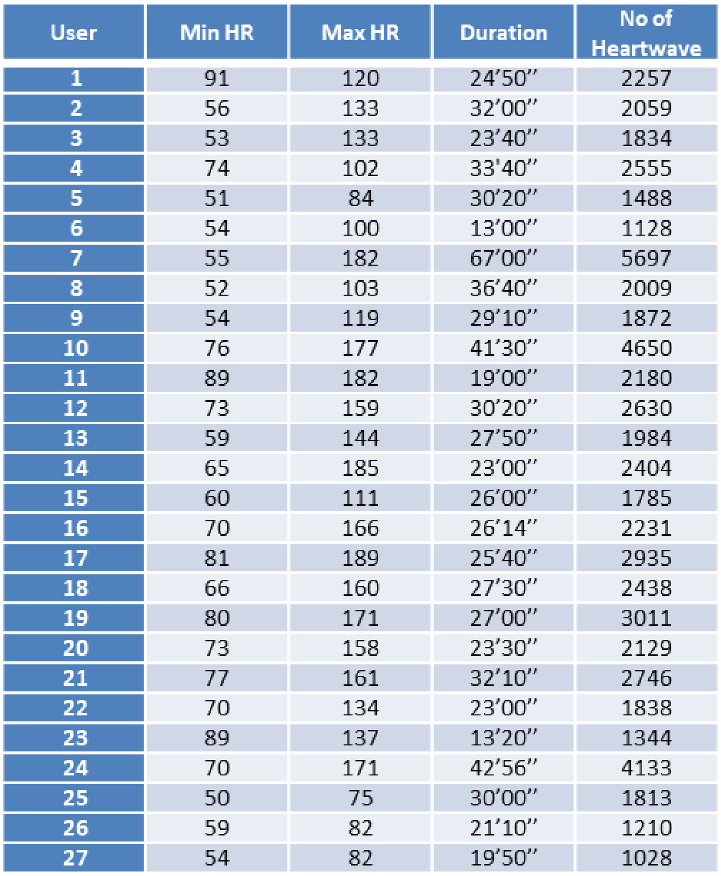
The validity of the detected features are evaluated using performance indicator Positive Predictive Value (PPV) of the following form:

$$PPV = \frac{No.of\ True\ Positive}{No.\ of\ True\ Positive + No.\ of\ False\ Positive} \tag{3.20}$$

where True Positive refers to the ability to detect the correct R peaks in accordance to the gold standard (reference to annotated R peak) and False Positive refers to an invalidated detection under gold standard (reference to annotated R peak). The Positive Predictive Value for R peak detection achieved 100%.

For the remaining 10 features, unfortunately, annotation from the dataset is not available. To determine the accuracy of the detection, 100 random heartwaves from each individual are selected and each of the features has been manually determined to the best ability from the author. In addition, a random selection of 10 users together with random selection of 20 heartwaves from each user have been positively validated by medical cardiologist. The accuracy of the detection for the remaining 10 features an average of 98.5%. Results of the accuracy is shown in Figure 3.21.

Table 3.1: Summary of heartwave signals from 27 users



User	Min HR	Max HR	Duration	No of Heartwave
1	91	120	24’50”	2257
2	56	133	32’00”	2059
3	53	133	23’40”	1834
4	74	102	33'40”	2555
5	51	84	30’20”	1488
6	54	100	13’00”	1128
7	55	182	67’00”	5697
8	52	103	36’40”	2009
9	54	119	29’10”	1872
10	76	177	41’30”	4650
11	89	182	19’00”	2180
12	73	159	30’20”	2630
13	59	144	27’50”	1984
14	65	185	23’00”	2404
15	60	111	26’00”	1785
16	70	166	26’14”	2231
17	81	189	25’40”	2935
18	66	160	27’30”	2438
19	80	171	27’00”	3011
20	73	158	23’30”	2129
21	77	161	32’10”	2746
22	70	134	23’00”	1838
23	89	137	13’20”	1344
24	70	171	42’56”	4133
25	50	75	30’00”	1813
26	59	82	21’10”	1210
27	54	82	19’50”	1028

In comparison, the performance from [59] that use combined Daubechies 4 and 6 wavelets solely, achieved an average of 98% accuracy. Importantly it is to the note that the compared method has only been tested on databases under resting heartrate and no compensation approach has been used to address QT and PR variability. When heartrate variability due to exercise duress is included, the accuracy drops to below 90%.

In another comparison, [119] use Laser Doppler Vibrometry methodology for pulse signal recognition and achieved an EER of 2.8%. However, the mentioned method suffered deteriorating performance when the heartrate exceeds 55% of individual maximum heartrate.

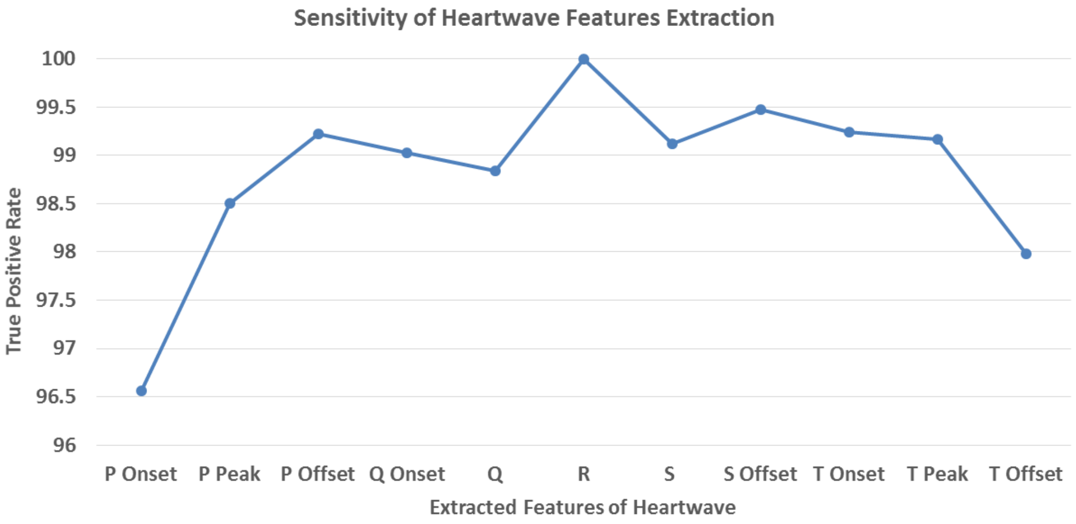


Figure 3.21: Accuracy of extracted features

3.4 Summary

The development of the heartrate adaptive extraction method for heartwave characteristic features has achieved a consistent accuracy and reliability. More importantly, the adaptive algorithm has the tenacity to tolerate morphological changes of the heartwave signal due to increasing heart rate. Equally important, algorithm provided a detection feature that includes the detection of onset and offset of QRS-Wave complex which has been unexplored previously. The extracted features will be used for data classification to support authentication which will be covered in Chapter 4.

CHAPTER 4

GAUSSIAN MIXTURE MODEL-HIDEN MARKOV MODEL WITH USER SPECIFIC THRESHOLD AND HEARTATE AS CRITERIA

4.1 Introduction

With the concluded extraction of heartwave features as described in Chapter 3, this chapter will describe the use of GMM-HMM methodology to classify the feature data to support user authentication. Chapter 4 commences with the presentation of the proposed architecture which includes the mathematical concepts of Gaussian Mixture Model and Hidden Markov Model. Thereafter, the results of the proposed architecture is presented and discussed.

4.1.1 Motivations

- To investigate the use of statistical methodology namely the Gaussian Mixture Modelling–Hidden Markov Modelling to support heartwave based classification under intense morphological changes of the heartwave signal.
- To investigate the effectiveness in integrating user unique parameters of user thresholding score and heartrate as part of the classification criteria to improve the performance of heartwave based classification.

4.2 Proposed architecture of GMM-HMM Topology

To cater for heartwave morphological variations, the proposed architecture uses a combined Gaussian Mixture Modelling with Hidden Markov Modelling (GMM-HMM) methodology. The GMM module of the architecture uses the extracted features to generate a GMM model for each of the users. To recap, individual heartrate can vary as much as 400% under intense physical exercise which causes significant heartwave morphological variation. The use of GMM on individual allows all variations of heartwaves to be contained in a single GMM model. Leveraging on the individual GMM model, the output of GMM model is further used to develop a classification model via the Hidden Markov Modelling module. This resulted into each user having a unique HMM model. To further strengthen the performance of the classification, 2 additional parameters namely, the individual threshold based on individual loglikelihood score together with individual heartrate range are incorporated in the classification. These parameters are individual dependent and unique to individual. Important, the incorporation of the individual dependent parameters addresses morphological variations of the heartwave signal to achieve a reliable heartwave based biometric authentication. See Figure 4.1 for a schematic summary of the proposed architecture.

The proposed methodology comprises of 3 stages. The first stage focuses on the data preparation. As the heartwave signals involved consist of users whose signals are acquired under intense physical duress and to ensure a reliable classification, repeatability and accuracy of the input data play an important role. The second stage focuses on the representation of data by Gaussian distribution. The joint distribution of the data is based on individual dataset and it comprises of heartwaves at different heartrate. The distribution model of the data via GMM allows a model representative of individual to support the subsequent classification. The last stage focuses on the generation of classification model using Hidden Markov Modelling. The classification also includes the use of user unique parameters to support classification.

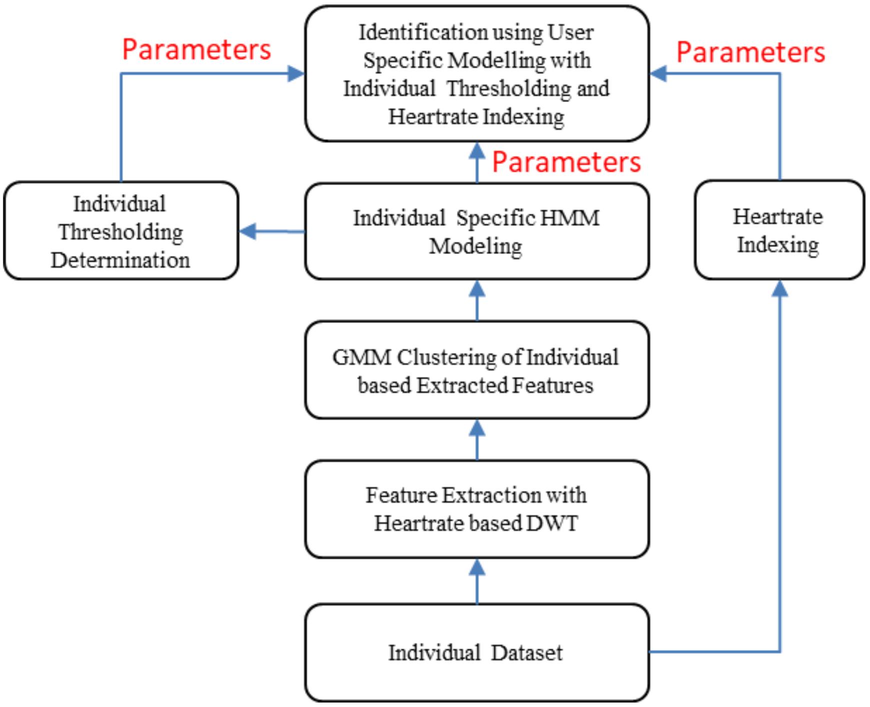


Figure 4.1: Proposed Architecture for GMM-HMM

4.2.1 Stage 1: Data Preparation

Chapter 3 reveals the extraction methodology of heartwave features. For each heartwave, a total of 11 unique features are extracted. To maintain consistency in the classification development, only Lead I and Lead II signals will be used for the development of classification algorithm. In the data preparation, the data from each individual is segmented into sequences of 10 seconds. Each of the sequences can contain from 8 to as much as 30 individual segmented sequential heartwaves which is proportional to the heartrate acquired at that instance. Please see Figure 4.2 for illustration of data preparation. In each of the heartwave, it contains 11 features as shown in Figure 4.3.

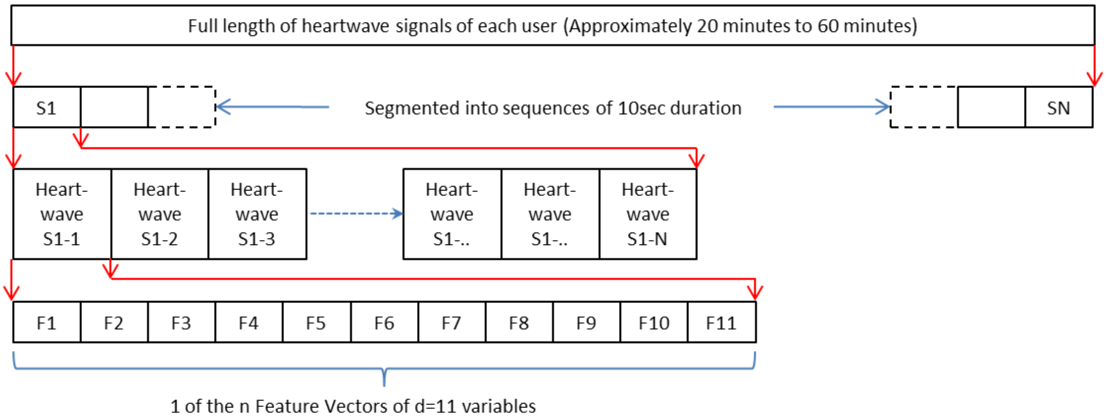


Figure 4.2: Organization of data to support GMM-HMM classification

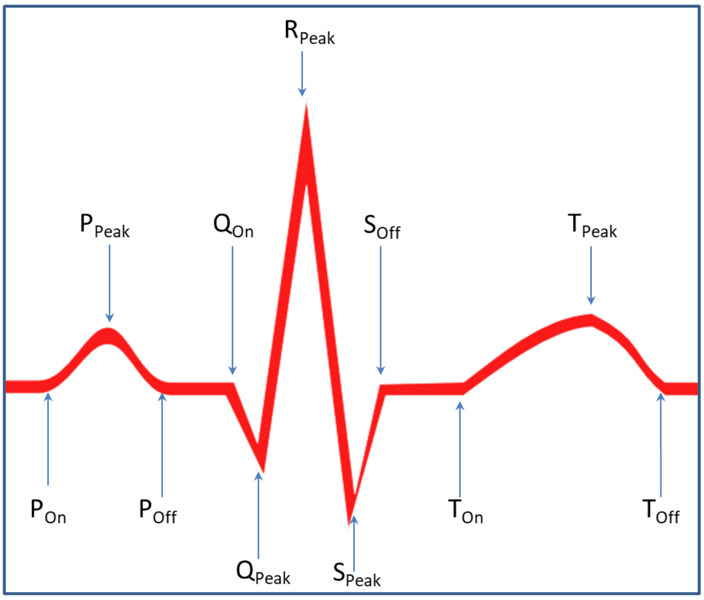


Figure 4.3: 11 features of a heartwave signal

Each sequence of heartwave data from an individual can comprise of signal morphologies from different states of the heartrate. During the acquisition process of heartwave signal from individual, an individual will start off in a resting state. The physical duress during the treadmill session will stress an individual physically at increasing intensity till the individual cannot tolerate the activity or when an individual has reached it maximum heartrate which is defined by the general guidelines of 240 - Age of Individual. Typically, the duration of maximum intensity lasts for few minutes.

Thereafter, the individual will undergo a recovery phase till the heartrate returns to resting state. Hence, during the short period of 10 seconds in a sequences, period of R-R interval is varying in particular to sequence acquired under intense physical duress.

During the data extraction, it is also observed that although heartrate has returned to the initial resting heartrate, for some individuals, the recovered heartwave are still different from the initial resting heartwave morphology. This implies that for some individuals, a longer recovery period of approximately 30mins is required before the heartwave can return to normal resting state. See Figure 4.4 for the variations of heartwave signal. Figure 4.4 shows a compilation of all the heartwave signals from the same individual at about R-Peak. The signal contains heartwave from resting heartrate to accelerated heartrate. Notice the variation of the waveforms in particular to the P-Wave and T-Wave. Hence with the understanding of the heartwave morphological changes in a single session of treadmill exercise, it can be hypothesized that each heartwave comprising of 11 fiducial parameters can be normally distributed into different components signifying different states of heartrates.

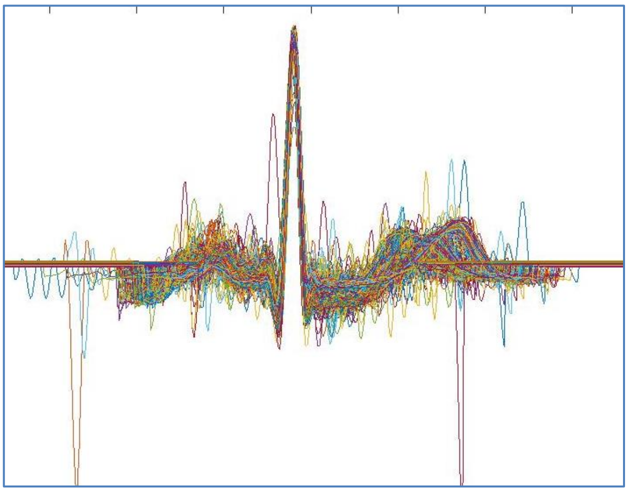


Figure 4.4: Superposition of all heartwaves from an individual

In total, there are 27 records of heartwaves signals acquired under treadmill testing. See Table 4.1 for details. The duration of the treadmill testing ranges from 13 minutes to as long as 67 minutes. Importantly, the minimum heartrate which is the resting heartrate varies from 50 beats per minute (bpm) to as high as 90 bpm and the maximum heartrate ranges from 82 bpm to as high as 190 bpm. The extremities that an individual can stretch from their resting heartrates to maximum heartrates ranges from 150% to more than 300%.

Table 4.1: Profiles of individual heartrate under treadmill testing

User	Min HR	Max HR	Duration	No of Heartwave
1	91	120	24’50”	2257
2	56	133	32’00”	2059
3	53	133	23’40”	1834
4	74	102	33'40”	2555
5	51	84	30’20”	1488
6	54	100	13’00”	1128
7	55	182	67’00”	5697
8	52	103	36’40”	2009
9	54	119	29’10”	1872
10	76	177	41’30”	4650
11	89	182	19’00”	2180
12	73	159	30’20”	2630
13	59	144	27’50”	1984
14	65	185	23’00”	2404
15	60	111	26’00”	1785
16	70	166	26’14”	2231
17	81	189	25’40”	2935
18	66	160	27’30”	2438
19	80	171	27’00”	3011
20	73	158	23’30”	2129
21	77	161	32’10”	2746
22	70	134	23’00”	1838
23	89	137	13’20”	1344
24	70	171	42’56”	4133
25	50	75	30’00”	1813
26	59	82	21’10”	1210
27	54	82	19’50”	1028

4.2.2 Gaussian Mixture Modelling (GMM)

4.2.2.1 Pre-GMM Data Preparation

To enable heartwave signal data to be represented by multiple Gaussian densities, it is necessary to prepare the data. This preparation is different from the preparation described in 4.2.1. The latter is the preparation needed for classification.

After implementing the heartwave delineation methodology, each of the heartwave sample consisting of 11 features is labelled as a sample vector. Theereafter all the individual heartwave signal are concatenated into a dataset of [N x D] where N is the total number of samples in individual and D is the 11 dimensions of heartwave features. See Figure 4.5 for illustration of the data preparation.

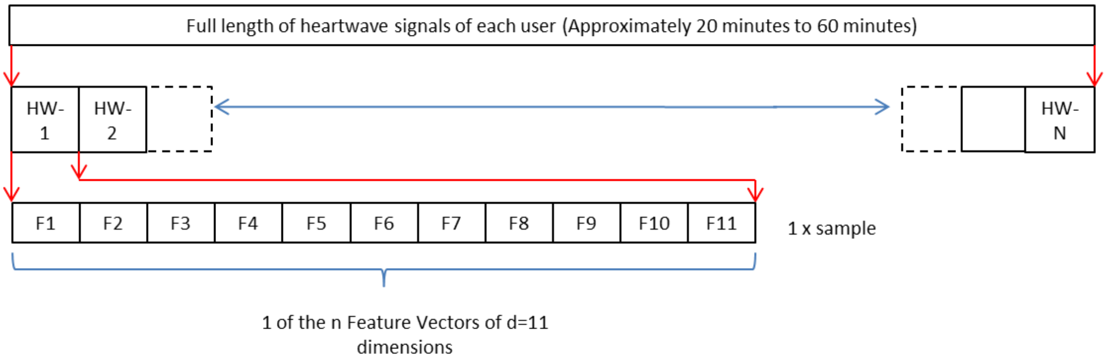


Figure 4.5: Data preparation for GMM

4.2.2.2 GMM Modelling

The concept of GMM methodology starts from the understanding of the Gaussian distribution, joint and conditional probability to identify the posterior probability or responsibility to best explain the observed sample. To achieve it, it rides on the well-established approach of Expectation Maximization (EM) to determine the parameters.

Gaussian mixture distribution which comprises of multiple Gaussian can be expressed as a linear superposition of Gaussian which is of the form

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \qquad (4.1)$$

$$\text{where} \quad \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_{ik})}$$

$p(x)$ refers to the probability density function of a mixture model, $\mathcal{N}(x|\mu_k, \Sigma_k)$ as the Gaussian distribution and π_k as the weightage also known as the prior probability or mixing coefficient. A K-dimensional binary random variable z have a 1-of-K representation is introduced. In the variable **z**, only a particular element z_k is equal to 1 and all the other elements are zero such that $\sum_k z_k = 1$ and $z_k \in \{0,1\}$. Therefore the distribution over the variable z with respect to the mixing coefficients π_k, is of the following:

$$p(z_k = 1) = \pi_k \tag{4.2}$$

where $\{\pi_k\}$ is satisfied by

$$0 \leq \pi_k \leq 1$$

$$\textstyle\sum_{k=1}^{K} \pi_k = 1$$

Expression (4.2) can be formatted into the distribution form and expressed as the following:

$$p(\mathbf{z}) = \textstyle\prod_{k=1}^{K} \pi_k^{z_k} \tag{4.3}$$

Expressing (4.3) as a Gaussian, the conditional distribution of **x** given a latent value for variable **z** is as follows:

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)^{z_k} \tag{4.4}$$

which can be expressed as

$$p(\mathbf{x}|\mathbf{z}) = \textstyle\prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)^{z_k} \tag{4.5}$$

The joint distribution of $p(\mathbf{x}, \mathbf{z})$ is $p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$. With the expressions of (4.3) and (4.4), the marginal distribution of **x** is the probability distribution of **x** averaging over the information about $\mathbf{z}$. This is calculated by summing the joint probability distribution over all possible states of $\mathbf{z}$ and is represented by the following:

$$\begin{aligned} p(\mathbf{x}) &= \textstyle\sum_{z} p(\mathbf{z})\, p(\mathbf{x}|\mathbf{z}) \\ &= \textstyle\sum_{k=1}^{\mathrm{K}} \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) \end{aligned} \tag{4.6}$$

Hence the marginal distribution of $\mathbf{x}$ is a Gaussian mixture of the form expressed in (4.1). For several observations $\mathbf{x_{1,} x_2 \ldots, x_n}$ the marginal distribution $p(\mathbf{x})$ of (4.6) can be re-expressed as $p(\mathbf{x}) = \sum_{z} p(\mathbf{x}, \mathbf{z})$ as for every observed data $\mathbf{x_n}$, there is a corresponding latent variable $\mathbf{z_n}$.

Next, we define a conditional probability of z given x. Let $\gamma(z_k)$ to denote $p(z_k = 1|\mathbf{x})$ where the value can be achieved by Bayes' theorem.

$$\begin{aligned} \gamma(z_k) &\equiv p(z_k = 1|\mathbf{x}) \\ &= \frac{p(\mathbf{x}|z_k = 1)p(z_k = 1)}{p(\mathbf{x})} \\ &= \frac{p(\mathbf{x}|z_k = 1)p(z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} \\ &= \frac{\pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\mu_j, \Sigma_j)} \end{aligned} \tag{4.7}$$

From expression (4.7), π_k is the prior probability of $z_k = 1$ with $\gamma(z_k)$ as the corresponding posterior probability based on the observed x. $\gamma(z_k)$ can also be viewed as the responsibility that component k takes for explaining the observation $\mathbf{x}$.

Given a dataset of observation $\mathbf{X} = \{\mathrm{x_{1,} x_2 \ldots, x_n}\}$ with the intention to model the data by Gaussian mixture, the maximum loglikelihood function can be applied to expression (4.1) through the application of log as shown.

$$p(\mathrm{x}|\pi,\mu,\Sigma)=\sum_{k=1}^{\mathrm{K}}\pi_k\mathcal{N}(\mathrm{x}|\mu_k,\Sigma_k)$$

$$\ln p(\mathbf{X}|\pi,\mu,\Sigma)=\sum_{n=1}^{N}\ln\left\{\sum_{k=1}^{\mathrm{K}}\pi_k\mathcal{N}(\mathrm{x}_n|\mu_k,\Sigma_k)\right\} \qquad (4.8)$$

Maximizing the loglikelihood expression of (4.8) is complex as it is not a close form due to presence of the summation over k that appears inside the logarithm. To maximize the loglikelihood, expression (4.8) can be solved by Expectation Maximization (EM). Prior to EM, to maximize the parameters, partial derivative is applied to (4.8) with respect to the parameters and setting the partial derivative to zero.

Optimization of the mean μ is achieved by performing a partial derivative of (4.8) with respect to the mean and setting it to zero.

$$\frac{\partial}{\partial\mu_k}\left(lnp(\boldsymbol{X}|\pi,\mu,\Sigma)\right)=0$$

$$=\frac{\partial}{\partial\mu}\left(\sum_{n=1}^{N}\ln\left\{\sum_{k=1}^{\mathrm{K}}\pi_k\mathcal{N}(\mathrm{x}_n|\mu_k,\Sigma_k)\right\}\right)$$

$$=\sum_{n=1}^{N}\frac{1}{\sum_j\pi_j\mathcal{N}(\mathrm{x}_n|\mu_j,\Sigma_j)}\frac{\partial}{\partial\mu}\{\pi_k\mathcal{N}(\mathrm{x}_n|\mu_k,\Sigma_k)\}$$

$$=\sum_{n=1}^{N}\frac{\{\pi_k\mathcal{N}(\mathrm{x}|\mu_k,\Sigma_k)\}}{\sum_j\pi_j\mathcal{N}(\mathrm{x}_n|\mu_j,\Sigma_j)}{\Sigma_k}^{-1}(\mathrm{x}_n-\mu_k)$$

$$=\sum_{n=1}^{N}\gamma(z_k)(\mathrm{x}_n-\mu_k) \qquad (4.9)$$

Hence from expression (4.9) it is now possible to determine μ_k and can be rearranged as follows:

$$\sum_{n=1}^{N}\gamma(z_k)(\mathrm{x}_n-\mu_k)=0$$

$$\sum_{n=1}^{N}\gamma(z_k)\,\mathrm{x}_n=\sum_{n=1}^{N}\gamma(z_k)\,\mu_k$$

$$\mu_k=\frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_k)\,\mathrm{x}_n \qquad (4.10)$$

$$where\ N_k = \sum_{n=1}^{N} \gamma(z_k)$$

The number of points assigned to a specific Gaussian component *k*, is the term, N_k. Expression (4.10) implies that μ_k for the *k^th^* Gaussian component is the weighted means of all the points in the dataset. Data point, x_n is factored with a weighting factor of posterior probability, $\gamma(z_k)$ from (4.7) that implies the likelihood of *k* generating the data, x_n.

Optimization of the covariance Σ_k similar to optimization of μ_k, through the partial derivative of (4.8) with respect to Σ_k and setting it to zero.

$$\frac{\partial}{\partial \Sigma_k}\left(lnp(\boldsymbol{X}|\pi,\mu,\Sigma)\right) = 0$$

$$= \frac{\partial}{\partial \Sigma_k}\left(\sum_{n=1}^{N} \ln\left\{\sum_{k=1}^{\mathrm{K}} \pi_k \mathcal{N}(\mathrm{x}_n|\mu_k,\Sigma_k)\right\}\right)$$

$$= \sum_{n=1}^{N} \frac{1}{\sum_j \pi_j \mathcal{N}\left(\mathrm{x}_n\middle|\mu_j,\Sigma_j\right)} \frac{\partial}{\partial \Sigma_k}\{\pi_k \mathcal{N}(\mathrm{x}_n|\mu_k,\Sigma_k)\}$$

$$\Sigma_k = \frac{1}{N_k}\textstyle\sum_{n=1}^{N} \gamma(z_k)\,(\mathrm{x}_n - \mu_k)(\mathrm{x}_n - \mu_k)^{\mathrm{T}} \qquad (4.11)$$

$$\text{where } \frac{\partial(\mathrm{x}-\mu_k)^T \Sigma_k^{-1}(\mathrm{x}-\mu_{ik})}{\partial \Sigma_k} = \Sigma_k^{-1}(\mathrm{x}_n - \mu_k)(\mathrm{x}_n - \mu_k)^{\mathrm{T}} \Sigma_k^{-1}$$

Similarly to maximization of Σ_k and μ_k, the mixing coefficients π_k, are also optimized and setting the partial derivatives to 0. However, to be coherent to expression $\sum_{k=1}^{K} \pi_k = 1$, a Larange multiplier is applied.

$$\frac{\partial}{\partial \pi_k}\left(\ln p(\boldsymbol{X}|\pi,\mu,\Sigma) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)\right) = 0$$

$$= \frac{\partial}{\partial \pi_k}\left(\sum_{n=1}^{N} \ln\left\{\sum_{k=1}^{\mathrm{K}} \pi_k \mathcal{N}(\mathrm{x}_n|\mu_k,\Sigma_k)\right\} + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)\right)$$

$$\frac{\partial}{\partial \pi_k}\left(\sum_{n=1}^{N}\sum_{k=1}^{\mathrm{K}}\left(-\frac{\gamma(z_k)}{2}\ln 2\pi_k - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}\sum_{n=1}^{N}(\mathrm{x}_n - \mu_k)(\mathrm{x}_n - \mu_k)^{\mathrm{T}}\Sigma_k^{-1}\right)\right.$$

$$\left. + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)\right)\Bigg) = 0$$

$$\frac{\partial}{\partial \pi_k}\left(\sum_{n=1}^{N}\sum_{k=1}^{\mathrm{K}}\left(\frac{D}{2}\ln 2\pi_k\right) - \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)\right) = 0$$

$$\sum_{n=1}^{N}\frac{\gamma(z_k)}{\pi_k} - \sum_{k=1}^{K}\lambda = 0$$

$$\sum_{n=1}^{N}\sum_{k=1}^{\mathrm{K}}\left(\frac{\gamma(z_k)}{\pi_k}\right) - \sum_{k=1}^{K}\lambda = 0$$

$$\sum_{n=1}^{N}\gamma(z_k) - \sum_{n=1}^{N}\sum_{k=1}^{K}\lambda\pi_k = 0$$

$$\sum_{n=1}^{N}\gamma(z_k) = N\pi_k$$

$$N_k = \mathrm{N}\pi_k$$

$$\pi_k = \frac{N_k}{N} \tag{4.12}$$

Expression (4.12) implies that the mixing coefficient for a particular component is the resulting average of the responsibility which the component takes for the representation of the data points.

With the expression of (4.10), (4.11) and (4.12), an iterative algorithm of EM can be applied to establish the optimum solutions to compute (4.7). It starts with an arbitrary assignment of values to the GMM parameters of mean, covariance and mixing coefficients including the initial loglikelihood score. EM algorithm comprises of 2 steps: Expectation (E-step) and Maximization (M-step). In E-step, based on the initial assignment of values into the parameters, the posterior probability based on expression (4.7) is evaluated. The evaluated new posterior probability is subjected to parameters maximization in M-step. In M-step, expressions (4.10), (4.11) and (4.12) are evaluated based on the new posterior probability. Thereafter the loglikelihood score from expression (4.8) is computed with new values of the evaluated parameters. Similarly, the newly computed parameters are input to re-compute the posterior probability. An iteration cycle of E-step and M-step increase the loglikelihood function. The solution is deemed to have converged when the difference of present and prior loglikelihood scores satisfies a user defined threshold or tolerance. Refer to the Figure 4.6 for flowchart of the EM algorithm.

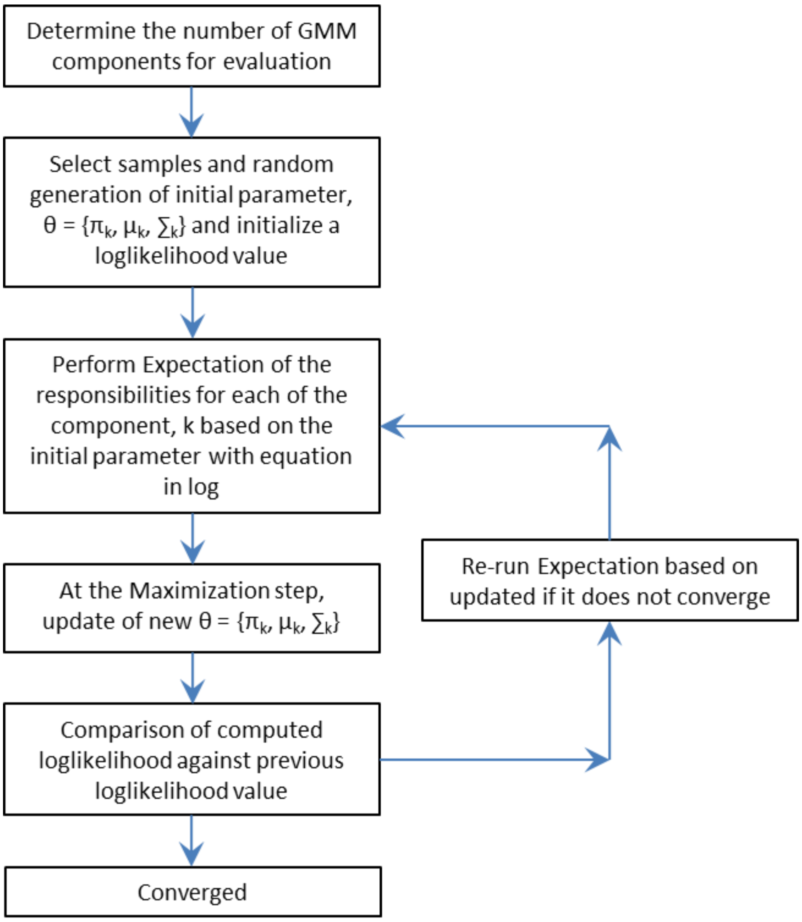


Figure 4.6: GMM EM for each user

With over 2000 heartwave signals from each user and more importantly, a complete heartwave morphology of an individual, GMM is used to represent each of the heartwave dataset. This results into each user having its unique parameters distribution.

In the development of the GMM for each user, it is known the higher the number of components in the GMM, the higher the loglikelihood score and it has the tendency to lead to infinity. Therefore in the GMM generation for each user, a further process to limit the number of components is executed. This work limits the model to be trained until over-fitting. As shown in Figure 4.7, as the number of components approaches infinitely, the loglikelihood result tends towards positive loglikelihood.

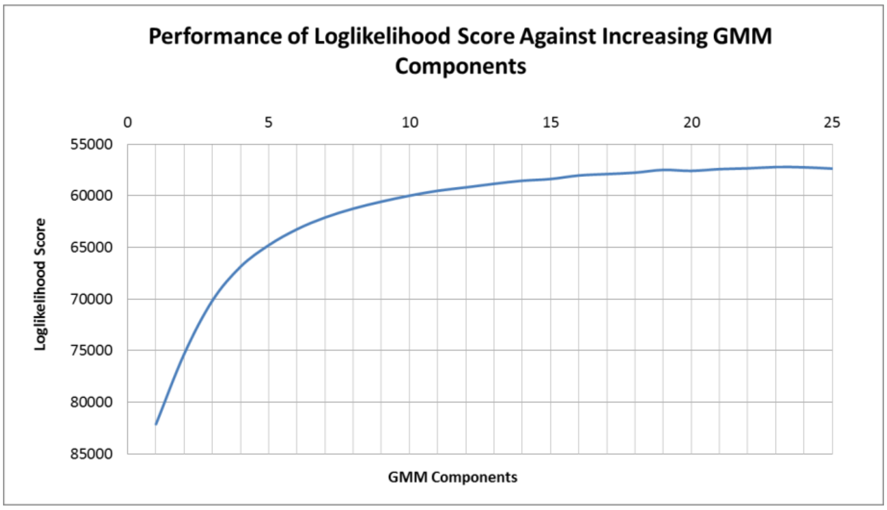


Figure 4.7: Increasing loglikelihood score against increasing components

Various approaches are available such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) or Minimum Description Length (MDL) to evaluate the most appropriate number of components for each user. Expressions (4.12), (4.13) and (4.14) for AIC, BIC and MDL respectively are various methods to implement a penalty term. AIC does not account for number of samples. As the size of the samples tends towards infinity, AIC estimate does not converge to true value since more data can potentially lead to an over-fitting model. AIC is typically used for unknown model with high dimensional reality. Both BIC and MDL are for true models as they account for the data and more explicitly, MDL accounts for total number of data values. This allows MDL to impose a heavier penalty for more complex model as compared to BIC estimation [120, 121]. These two approaches have been consistently used for model order determination:

$$AIC(K,\theta) = -2(logp(x|\mu,\Sigma,\pi)^{ML}) + 2L \tag{4.13}$$

$$BIC(K,\theta) = -2\sum_{i=1}^{N} \log p(x_i|\theta_i)^{ML} + (L)\log(N) \tag{4.14}$$

$$MDL = -\sum_{i=1}^{N} \log p(x_i|\theta_i)^{ML} + (\frac{L}{2})\log(dN) \tag{4.15}$$

where $= K\left(1 + d + \frac{d(d+1)}{2}\right) - 1$,

d is the number of features,

N is the number of data values,

K is the number of components.

In the determination on the limits of GMM components, the BIC and MDL methods are used. The loglikelihood results of BIC and MDL from each component are compared to the next higher components. When the existing loglikelihood from current component is lower than the loglikelihood from the next higher components, the current component is deemed the limited parameter.

Table 4.2: Component limit for each user in GMM modelling

User	Min HR	Max HR	No of Sequences	BIC	MDL
300	91	120	149	19	19
301	56	133	190	19	19
302	53	133	142	16	16
304	51	84	182	18	14
307	52	103	219	20	19
309	76	177	249	23	23
310	89	182	114	18	18
311	73	159	182	0	24
312	59	144	166	23	23
313	65	185	138	21	18
316	81	189	150	15	15
320	77	161	193	21	21
321	70	134	138	23	17
322	89	137	80	12	12
325	59	82	127	13	13
327	54	82	119	21	21

Table 4.2 shows the component limits for all the users. Most of the users have an optimum component at approximately 20. User_311 did not achieve optimum component using BIC Criterion unless the range of components testing is extended. Using MDL however established the optimum component at 25. The table also shows similar optimum components for both BIC Criterion and MDL for most users.

4.2.3 Hidden Markov Modelling

From the GMM process for each user, it is assumed that the optimized user specific GMM is sufficiently appropriate as a representative of the full data distribution for each individual. Hence, the number of ‘k’ components for each user is equivalent to the number of states for Hidden Markov Modelling (HMM). The optimized parameters values of GMM are input to the initial states of the HMM parameters. In order for the HMM to be used as classification model, similar EM is employed to optimize the parameters of the HMM model.

Base on the understanding of the heartwave signal variations for each user, given a dataset of sequences, the heartwave can exist at any stationary state or any series of state sequence. For example, during the recovery from strenuous exercise, an individual heartrate can progress from a heartrate of 120bpm to a recovery heartrate of 90bpm within a span of 2 minutes. The use of HMM will be able to statistically determine the progression of states given a sequence or concatenated sequences. With this capability, classification of a heartwave sequence can be used to support user classification.

HMM can be presented as the simplest Bayesian Network and HMM has been widely use in speech recognition, gesture recognition and even medical diagnostic application. The latter has been widely used to detect heartwave anomaly. It is an approach for modelling dynamics systems that are observed through a time-series. HMM is characterized by the following 5 traits.

1. The number of N hidden states within the model. Each state corresponds to a unique state provided by the model.
2. The number of unique observations per state. These symbols are denoted as:
$$V = \{v_1, v_2, \ldots., v_M\}. \tag{4.16}$$
3. State transition probability distributions $A = \{a_{ij}\}$ where
$$a_{ij} = P\left(q_{t+1} = S_j \,|q_t = S_i\right), \;\; 1 \leq i, j \leq N \tag{4.17}$$
4. The emission probability distribution in state $j, B = \{b_j(k)\}$ where
$$b_j = P\left(v_k \; at \;\; t \,|q_t = S_j\right), \;\; 1 \leq j \; \leq N, 1 \leq k \; \leq M \tag{4.18}$$
5. The prior probability $\pi_i = \{\pi_i\}$ of being in state i at the beginning of the observations where
$$\pi_i = P(q_1 = S_i), 1 \; \leq i \; \leq N \tag{4.19}$$

In a simplified representation HMM methodology as shown in Figure 4.8, observations are generated according to the probability distribution known as the emission probability $P(O|S)$. The emission probabilities do not change over time. In addition, the state that generates the respective observation is only dependent on the previous state which is governed by transmission probability $P(S_2|S_1)$. . Hence, the observation is independent of all other variable except on the current state.

In the generation of HMM for each user, each model from each user will have a model parameter represented by $\boldsymbol{\lambda_{user,k}}$. The $\boldsymbol{\lambda_{user,k}}$ comprises of $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi})$ where **A** is the state transition probability distribution $\{a_{ij}\}$, **B** is the observation probability distribution $\{b_i(v_k)\}$ where v is the symbol observation at the respective state and **Π** as the initial state distribution. For the proposed architecture, for a given set of sequences, HMM is used to determine the probability or the likelihood of the observed sequence. With HMM model for each of the user, the classification is based on the HMM model that generates the highest probability. Prior to the classification, it is necessary to train and generate a HMM model for each user.

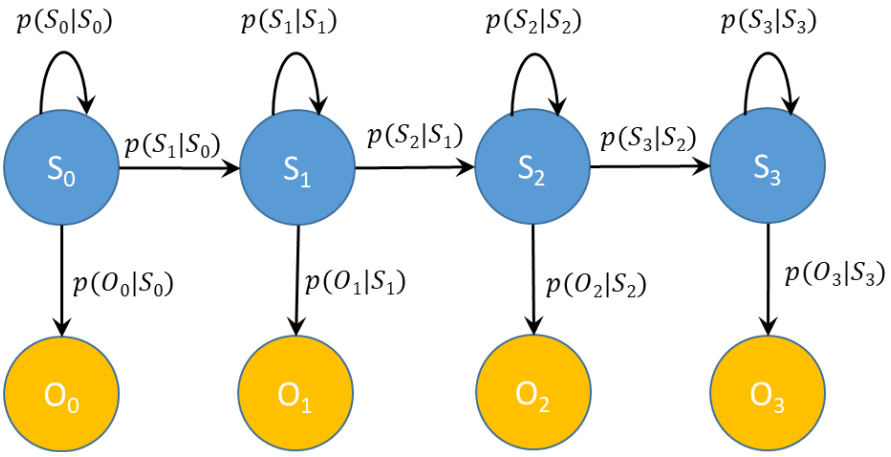


Figure 4.8: Illustration of a HMM model

Let $\lambda = (A, B, \pi)$ be the parameters of the HMM model and $O = (O_0, O_1, \dots, O_{T-1})$ be the series of observations from the sequence. The objective is to find $P(O|\lambda)$. $S = (s_0, s_1, \dots, s_{T-1})$ refers to the state sequence and hence the observation is as follow using the forward algorithm.

$$\alpha_t(i) = P(O_0, O_1 \dots O_T, s_t = q_i|\lambda) \qquad (4.20)$$

where $\alpha_i(i)$ is the probability of the partial observation sequence at state q_i at time *t*, for *t=0,1,…T-1* and number of states being *i=0,1,…N-1*. Recursively computation commences from the initial state where $\alpha_0(i) = \pi_i b_i(O_0)$ as the initial state when *t=0*. For *t=1,2,…T-1,* the probability of the partial observation at state *i*, time *t* is the summation of all previous states ending at state *i*, time *t* and is expressed as follows for every node:

$$\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j)\, a_{ji}\right] b_i(O_t) \qquad (4.21)$$

$$\text{for } t = 1,2, \ldots T-1 \text{ and } i = 0,1, \ldots N-1$$

Hence, the total probability at *T-1* is shown in expression (4.22). See Figure 4.9 for a graphical representation of the forward algorithm on a lattice.

$$P(O|\lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i) \qquad (4.22)$$

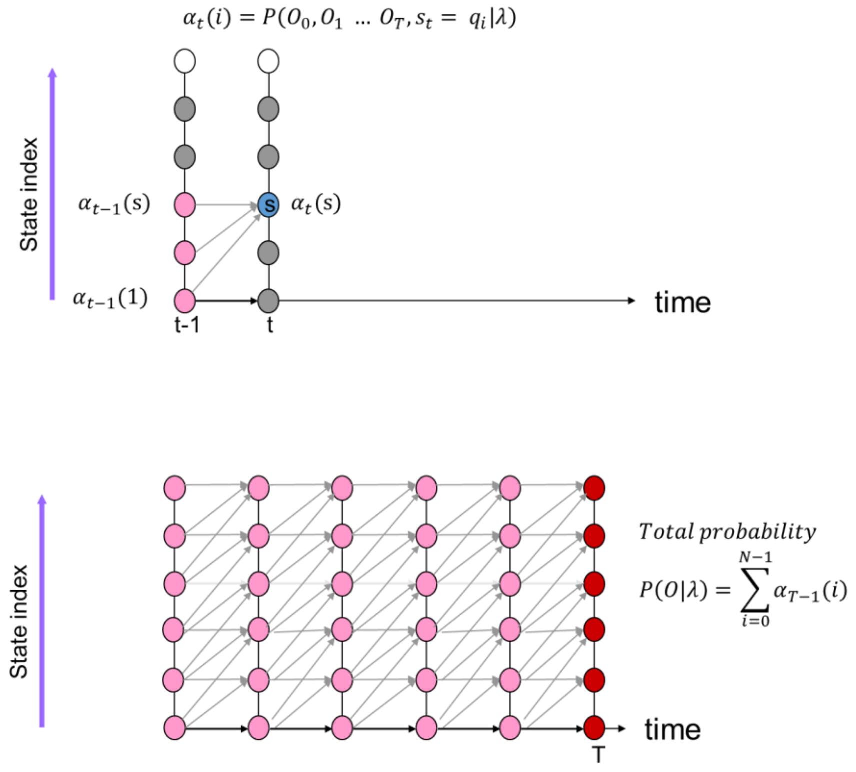


Figure 4.9: Forward algorithm on a trellis representation

Similar to GMM, HMM model needs to be train. Firstly, it is necessary to establish the posterior probability. Prior to establishing the posterior probability, it is also necessary to define the backward algorithm. The backward algorithm is defined by the following expression.

$$\beta_t \ (i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \, \beta_{(t+1)}(j) \tag{4.23}$$

$\beta_t(i)$ is defined as the total probability of all state sequences that depart from *s* at time *t* and all observations after q_t. See Figure 4.10 for a graphical representation of the backward algorithm on a lattice

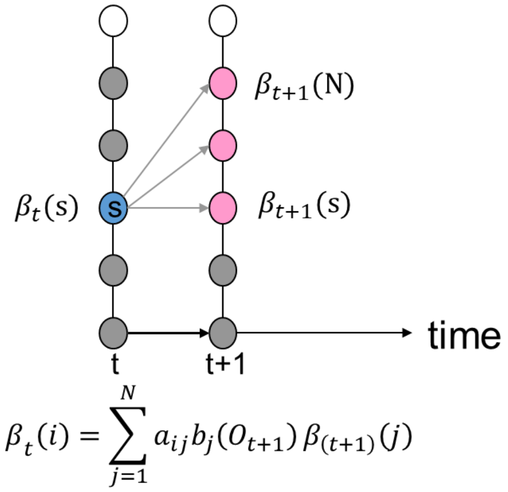


Figure 4.10: Representation of a backward algorithm

With the established expression of backward algorithm, the probability of the state can be determined. This is also known as the E-step of Expectation Maximization. It defines that the probability assigned to any state *s*, for any observation o_t is the probability that the process is at state *s* when it generates o_t.

$$p(state(t) = s | o_1, o_2, \dots o_T) \propto p(state(t) = s, o_1, o_2, \dots o_T) \tag{4.24}$$

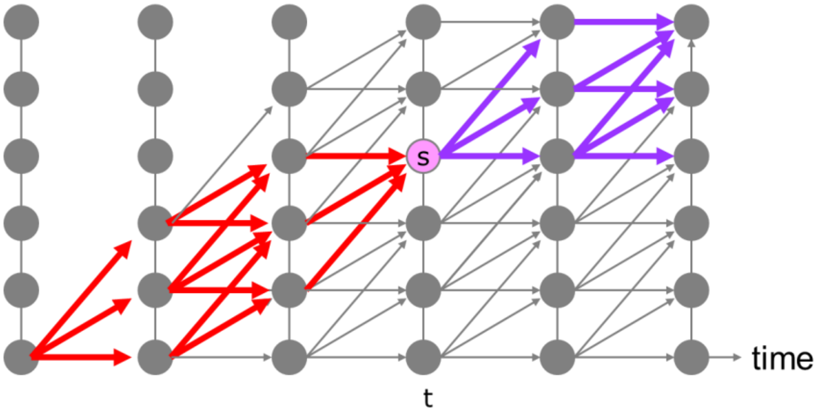


Figure 4.11: Lattice representation of Forward Backward algorithm

Therefore the probability of assigning an observation to a state comprises of the total probability the all state sequences ending at state *s* at time *t* and the total probability of all state sequences that begins at state *s* at time *t*. See Figure 4.11 and Figure 4.12 for illustration. The complete probability is shown in expression (4.25).

$$\alpha_t(\mathrm{s})\ \beta_t(\mathrm{s}) = p(state(t) = s, o_1,\ o_2, \dots o_T|\lambda) \tag{4.25}$$

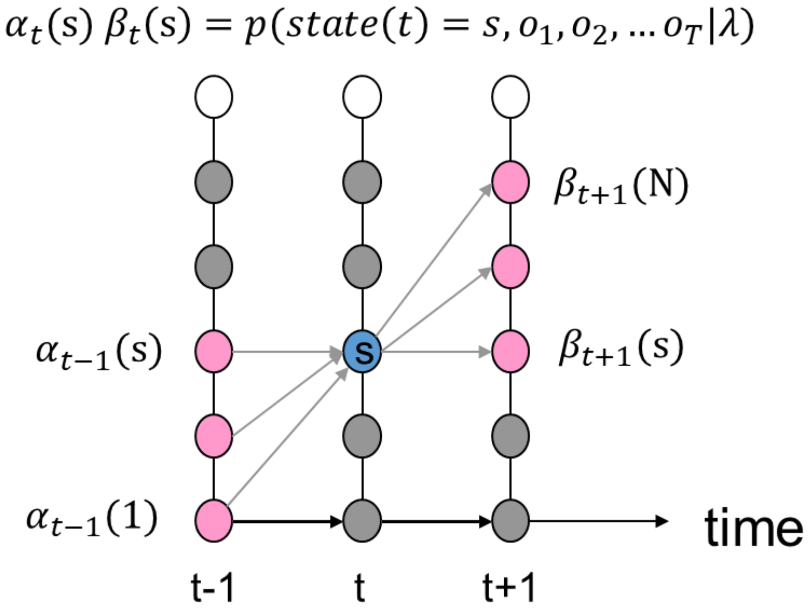


Figure 4.12: Complete probability of state *s* at time *t*

With the complete probability determined, the posterior probability can be achieved by normalization via the summation of all other states.

$$
\begin{aligned}
p(state(t) = s|O, \lambda) &= \frac{p(\boldsymbol{O}, state(t) = s|\lambda)}{\sum_{s'} p(\boldsymbol{O}, state(t) = s'|\lambda)} \\
&= \frac{\alpha_t(\mathrm{s})\beta_t(\mathrm{s})}{\sum_{s'} \alpha_t(s')\beta_t(s')} \\
&= \gamma_i(t) \qquad (4.26)
\end{aligned}
$$

Following, probability of the sequence at state *i* at time *t* and state *j* at time *t+1* is computed using

$$\xi_{ij}(t) = \frac{\alpha_i(t)a_{ij}b_j(O_{t+1})\beta_j(t+1)}{\sum_{i=1}^{N} \alpha_t(T)} \qquad (4.27)$$

In the M-step, the following parameters are updated and subsequently test for observation convergence

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_i(t)a_{ij}b_j(O_{t+1})\beta_j(t+1)}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)} \qquad (4.28)$$

$$\hat{\mu} = \frac{\sum_{t=1}^{T} \gamma_i(t)O_t}{\sum_{t=1}^{T} \gamma_i(t)} \qquad (4.29)$$

$$\Sigma_i = \frac{\sum_{t=1}^{T} \gamma_i(t)(O_t - \hat{\mu})(O_t - \hat{\mu})^T}{\sum_{t=1}^{T} \gamma_i(t)} \qquad (4.30)$$

The summary of the EM for HMM is summarized in the Figure 4.13.

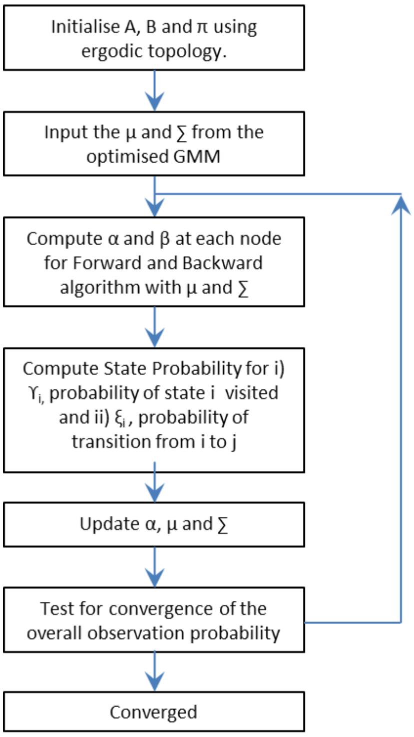


Figure 4.13: EM for HMM modelling

4.3 Experimentation and Results

4.3.1 Parameter Optimization: Size of Training Data

This investigation examines the percentage of data necessary to train a GMM-HMM model while ensuring the authentication will not be compromised. In actual implementation, it is not possible to train HMM model based on all data and at the same time to continuously train HMM upon an addition of new data. Hence this investigation uses the results of loglikelihood as a criteria to determine to most appropriate proportion of data for training.

The size of the training dataset starts off at a fixed percentage from 10% to 100%. Only sequences not using in the training of HMM parameters are used in the testing. In other words, data sequences that are used in the testing are not used for training of HMM model. The selection of the sequences for training is generated randomly to eliminate biasness.

The testing is conducted on all users and the raw results are tabulated. As each user has different number of sequences in their respective dataset, an appropriate approach is to use percentage as an incremental step. As evident from the table, using 10% to 20% of data sequences for training has a tendency to under-fit which leads to higher means and standard deviations. Conversely, having too much data for training will lead to a situation of over-fit as evident from Figure 4.14. Hence, a stable region will be within the range of 40% to 70% which can be appropriate for HMM parameter training. This work uses the upper bound of 70% for a more conservative approach. Figure 4.14 shows a normalized mean score based on the percentage of training data. The scores are normalized to 1 based on the 10% training data. Evidently, the score of individual is homogenous based within a range of 40% to 70% training data.

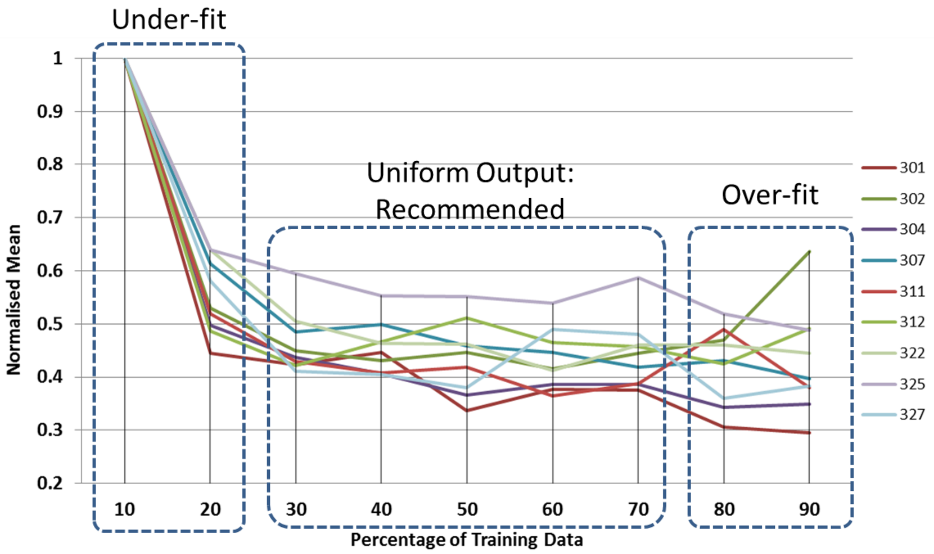


Figure 4.14: Distribution of normalized mean scores at increasing training data

4.3.2 Parameter Optimization: GMM-HMM State Optimization

Leveraging on the proposed 70% of data as training data, the remaining 30% sequences are used in the authentication investigation. Prior to authentication investigation, it is important to determine what is the optimum number of states required by the HMM module. The investigation is performed by collating all test sequences from all users. Each of the test sequences is tested on all the individual HMM model. For each test sequence, the HMM model that output the maximum loglikelihood result is indexed. See Figure 4.15 for the schematic of the investigating procedure.

To ensure non-biasness, the testing are carried out over 5 runs. In each of the run, the 30% testing data and 70% training data are re-randomized. The GMM-HMM model for each user are re-trained. The test sequences are re-subjected to testing as all over again. The results are accumulated, compiled and analyzed. 25 sub-models for each users are created and trained with 70% data. Simply, each user will have 25 HMM models ranging from HMM model with 1 state to model with 25 states.

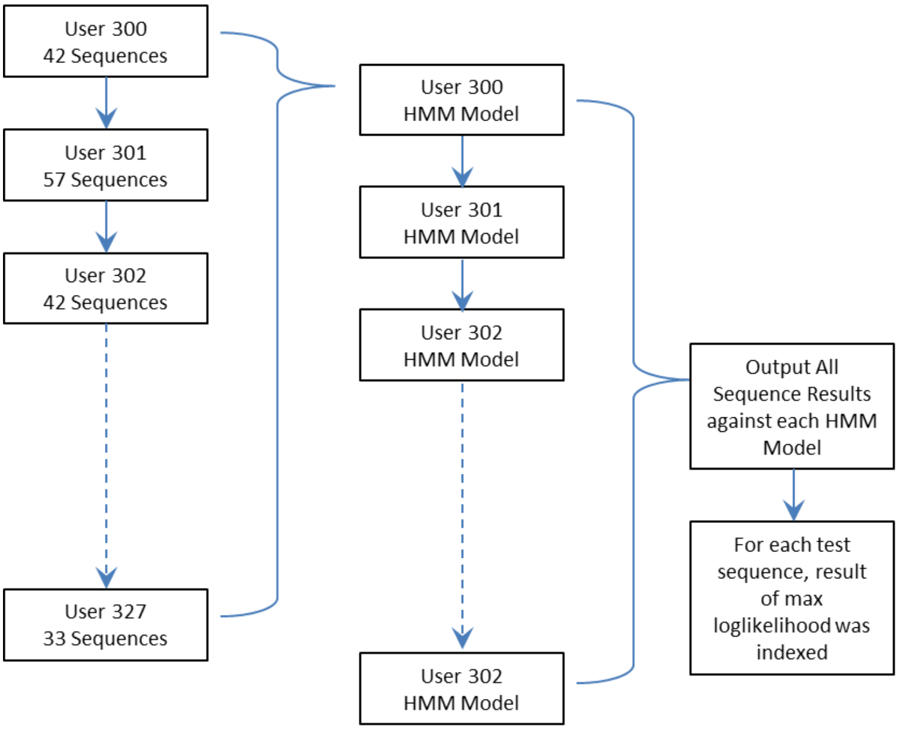


Figure 4.15: Testing of the test sequences against a database of 27 HMM model for identification testing

The results from the 5 runs are accumulated and tabulated in a confusion table as shown in Table 4.3. In the confusion table, the rows represent the actual class and the column represents the classification performance. The values along the major diagonal represents the correctly identified user with respect to the test sequence. Errors or values along the rows and between the major diagonal is known as the error of omission. It refers to the False Negative (FN). In an identification process, it means an authorized user has been denied access. Error or value along the vertical (below or above the major diagonal) is referred to as False Positive (FP). In identification process, it means unauthorized individual has been granted access to the system.

Table 4.3 shows the confusion table of the classification testing performed at State 1 together with the computational percentage of FP and FN. From the Confusion Table shown, User 312 has an Error of Commission at 16.68%. This anomaly is primarily due to the Error of Omission contributed by User 313 and User 320. Figure 4.16 shows the signal morphologies of User 312, 313 and 320. Comparing the signals of the mentioned User 312, 313 and 320, temporally the signals varies minimally. The deterioration of the accuracy in particular to User 312 can be attributed to the exclusion of amplitude profile as one of the features. This is evident from the varying R-Peak, Q-Wave Peak and even the T-Wave Peak among the three morphologies. Equally importantly, in the classification process that uses HMM as the basis, data sequence of heartwave signals in fixed duration of 10 seconds is used to determine the likelihood scores against the 27 HMM-GMM Models. The use of the data sequence averages the discriminating features as compared to using a singular heartwave signal.

In the testing and validation, each user has 25 sub-models of HMM that ranges from HMM model with 1 state to HMM model with 25 states. Hence, 25 such tables have been tabulated. From the tabulated confusion table at each state, the computed percentage of FP and FN data facilitates easy computation of parameters comprises of True Positive (TP), True Negative (TN), False Positive, False Negative, False Positive Rate (FPR) and True Negative Rate (TNR). Importantly, it facilities the computation of sensitivity and specificity which have been derived from TPR and TNR, respectively. See Table 4.4 for results.

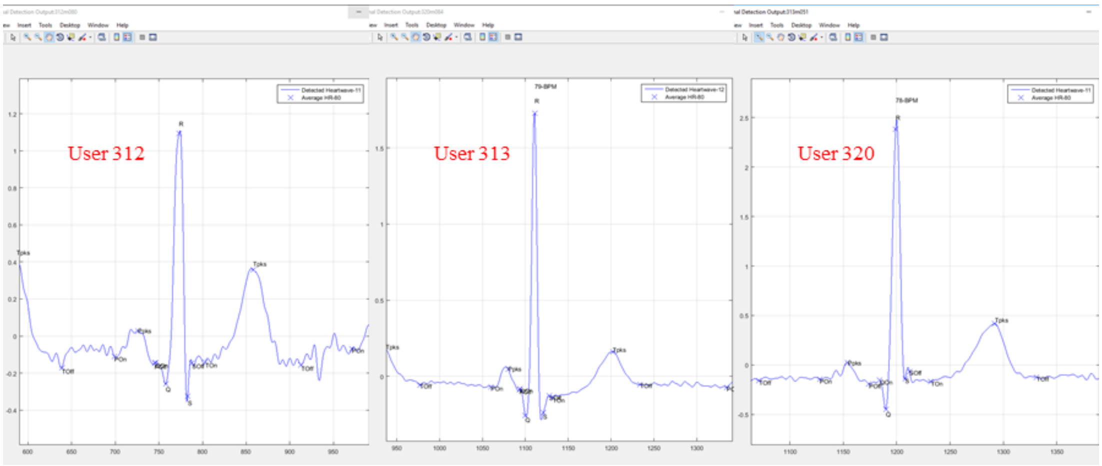


Figure 4.16: Heartwave profile for User 312, User 313 and User 320.

At each of the 25 sub-model state, the results are summed and analyzed based on the overall classification process involving 27 users. Figure 4.17a, 4.17b, 4.18a and 4.18b show the overall results for each of the individual states.

Figure 4.17a shows the Sensitivity analysis which is the ability to authorize access to the correct users. It shows the average sensitivity at different state. Figure 4.17b also shows that sensitivity achieved well over 90% at lower states between State 1 and State 5. This aligns well with the objectives that BIC and MDL criterion are useful to limit over-specification and conserve computation processing. It does not consider the effects of over-fit and under-fit.

Table 4.3: Confusion Table showing the accumulation of classification results from 5 runs at State 1

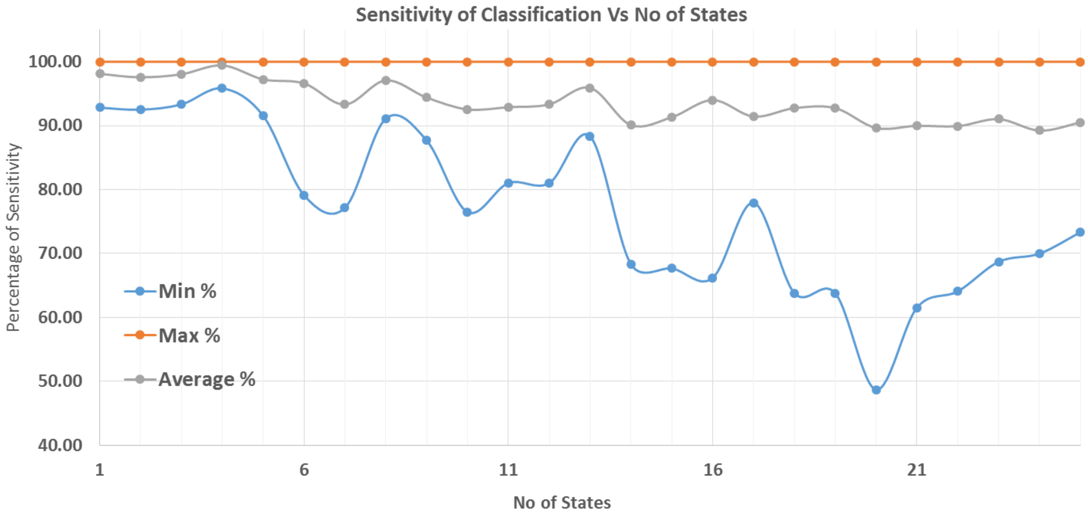
State 1																		
User	300	301	302	304	307	309	310	311	312	313	316	320	321	322	325	327	Total	Err Omis	% Omis
300	207	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	210	3	1.43
301	0	283	0	0	0	0	0	0	1	0	0	0	0	0	1	0	285	2	0.70
302	0	2	202	0	0	0	0	0	6	0	0	0	0	0	0	0	210	8	3.81
304	0	0	0	270	0	0	0	0	0	0	0	0	0	0	0	0	270	0	0.00
307	0	0	0	0	315	0	0	0	0	0	0	0	0	0	0	0	315	0	0.00
309	0	0	0	0	0	349	0	0	9	0	2	0	0	0	0	0	360	11	3.06
310	0	0	0	0	0	0	165	0	0	0	0	0	0	0	0	0	165	0	0.00
311	0	0	0	0	0	0	0	267	0	0	0	0	0	3	0	0	270	3	1.11
312	0	2	0	0	0	0	0	0	238	0	0	0	0	0	0	0	240	2	0.83
313	0	0	0	0	0	0	0	0	13	181	1	0	0	0	0	0	195	14	7.18
316	0	1	0	0	0	0	2	0	2	2	218	0	0	0	0	0	225	7	3.11
320	0	0	0	0	0	0	0	0	14	0	0	270	0	0	1	0	285	15	5.26
321	0	0	0	0	0	0	0	0	0	0	0	0	191	0	0	4	195	4	2.05
322	0	0	0	0	0	0	0	0	0	0	0	0	0	120	0	0	120	0	0.00
325	0	0	0	0	0	0	0	0	3	0	0	0	0	0	177	0	180	3	1.67
327	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	165	165	0	0.00
Total	207	291	202	270	315	349	167	267	286	183	221	270	191	123	179	169	3546	72	
Err Com	0	8	0	0	0	0	2	0	48	2	3	0	0	3	2	4	72		
% Com	0.00	2.75	0.00	0.00	0.00	0.00	1.20	0.00	16.78	1.09	1.36	0.00	0.00	2.44	1.12	2.37			

Table 4.4: Computation of biometric assessment parameters at State 1

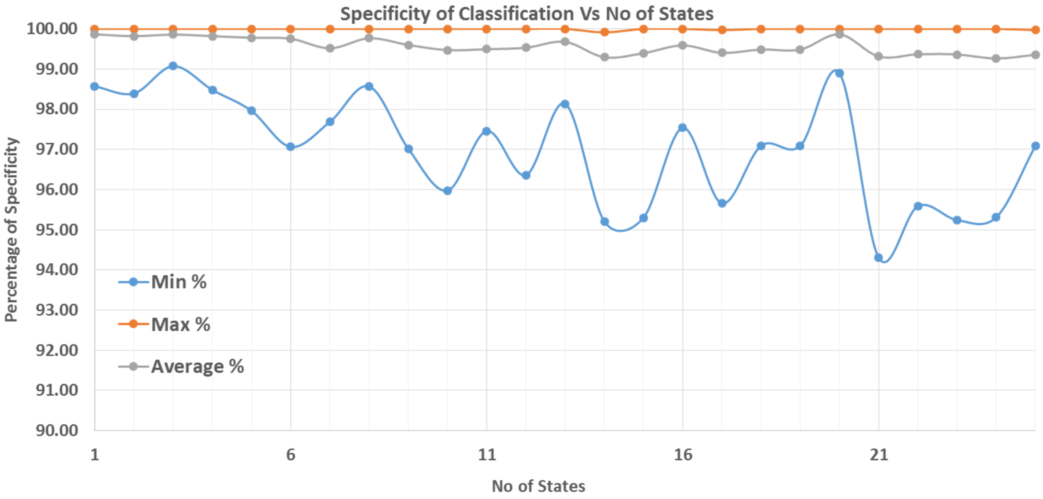
							Sensitivity	Specificity
User	**TP**	**FN**	**FP**	**TN**	**FPR**	**FNR**	**TPR**	**TNR**
300	207	3	0	3339	0.00	1.43	98.57	100.00
301	283	2	8	3263	0.24	0.70	99.30	99.76
302	202	8	0	3344	0.00	3.81	96.19	100.00
304	270	0	0	3276	0.00	0.00	100.00	100.00
307	315	0	0	3231	0.00	0.00	100.00	100.00
309	349	11	0	3197	0.00	3.06	96.94	100.00
310	165	0	2	3381	0.06	0.00	100.00	99.94
311	267	3	0	3279	0.00	1.11	98.89	100.00
312	238	2	48	3308	1.43	0.83	99.17	98.57
313	181	14	2	3365	0.06	7.18	92.82	99.94
316	218	7	3	3328	0.09	3.11	96.89	99.91
320	270	15	0	3276	0.00	5.26	94.74	100.00
321	191	4	0	3355	0.00	2.05	97.95	100.00
322	120	0	3	3426	0.09	0.00	100.00	99.91
325	177	3	2	3369	0.06	1.67	98.33	99.94
327	165	0	4	3381	0.12	0.00	100.00	99.88
Overall	**3618**	**72**	**72**	**53118**	**0.14**	**1.95**	**98.05**	**99.86**

Figure 4.17b shows the Specificity analysis which is defined as ability to deny access to unauthorized individuals. The results of the Specificity for all 25 states have been impressive as the specificity is well above 90% for all levels of states.

While the sensitivity and specificity have shown impressive results, the False Positive and False Negative in contrast provided avenues for more robust investigation. 4.18a shows the False Positive results at different states represented by 3 series of data which are the Percentage of Minimum, Maximum and Average at each stage. Minimum refers to the minimum value of False Positive among the group of 16 users at each State. The same definition applies to Maximum and Average.



(a)



(b)

Figure 4.17: HMM results for 25 different models: (a) Sensitivity. (b) Specificity.

At lower states of up to State 4, the False Positive is 20% and below. Beyond State 4, the False Positive Errors are in excess of 20% exclude State 8 and State 20.

In the False Negative results as shown in Figure 4.18b where authorized individuals are denied access, the False Negative errors are below 10% at State 5 and below. Based on best fit, the error worsen at increasing number of States. This observation is in-line with False Positive error and Sensitivity results that lower states provide better classification results. Thus this analysis concluded that at low number states up to 3 states, offers the most appropriate selection without compromising the results of the confusion table.

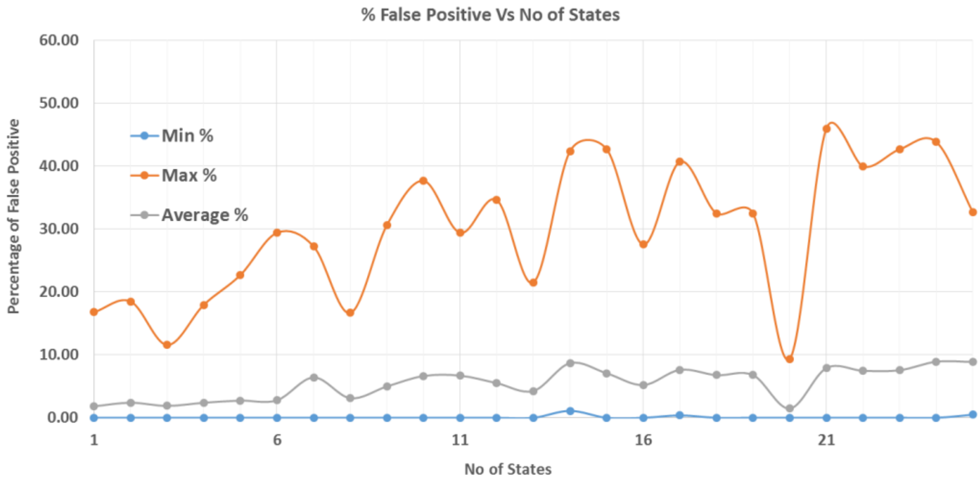


Figure 4.18a: Percentage of False Positive for 25 different HMM models

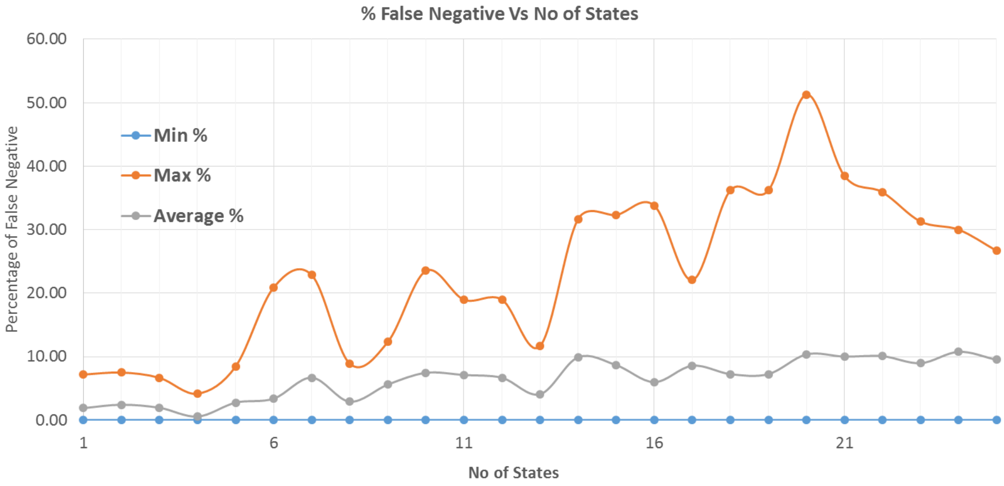


Figure 4.18b: Percentage of False Negative for 25 different HMM models

4.3.3 User Classification with User Specific Thresholding

Recap in Figure 4.17 and Figure 4.18, where the classification is based on the maximum likelihood score, it is not practical to use a model generated score to perform classification in particular when the population grows too big. An important observation is the presence of user specific parameters maximum and median loglikelihood scores. Figure 4.18 shows the distribution of the loglikelihood score from User316. Notice that the loglikelihood scores are homogenous at the heartrate of 120bpm and below. When the heartrate is between 140bpm to 180bpm, the loglikelihood scores are in the range between -600 to -1000. This is attributed to the morphological changes in the heartwave signal where the heart is under supraventricular tachycardia mode. Supraventricular tachycardia is a condition where the P-wave signal is eclipsed by T-wave of previous heartwave signal. For User316, the resting heartrate is around 80bpm. Hence, Figure 4.19 shows the stability of loglikelihood score at heartrate within the range of 80bpm to 140bpm but decreases for heartrate from 140bpm and above.

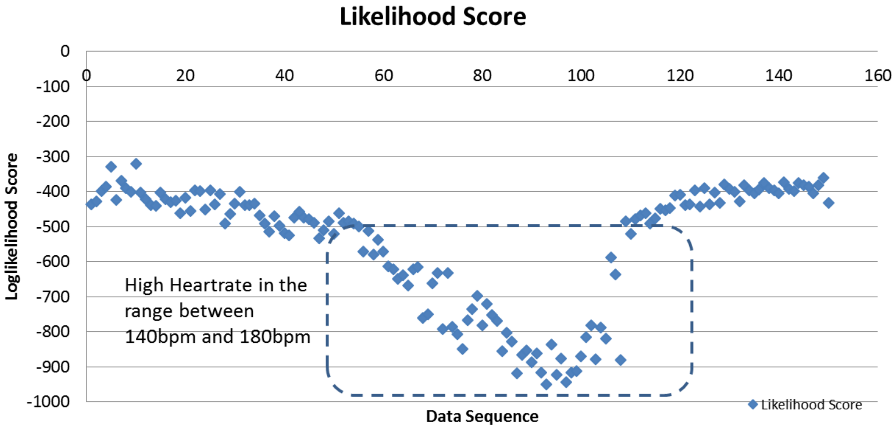


Figure 4.19: Varying loglikelihood score due to varying heartrate from User316

See Figure 4.20 for user specific score. Due to the nature of individual physiological signals, the user specific thresholding criteria on median loglikelihood score can be used as a user-specific thresholding criteria in classification.

In this validation, all individual’s 30% of the untrained data sequences are input as a lot to all the user specific GMM-HMM model to perform classification. The criteria to classify each sequence is dependent on user specific thresholding criteria of median loglikelihood score.

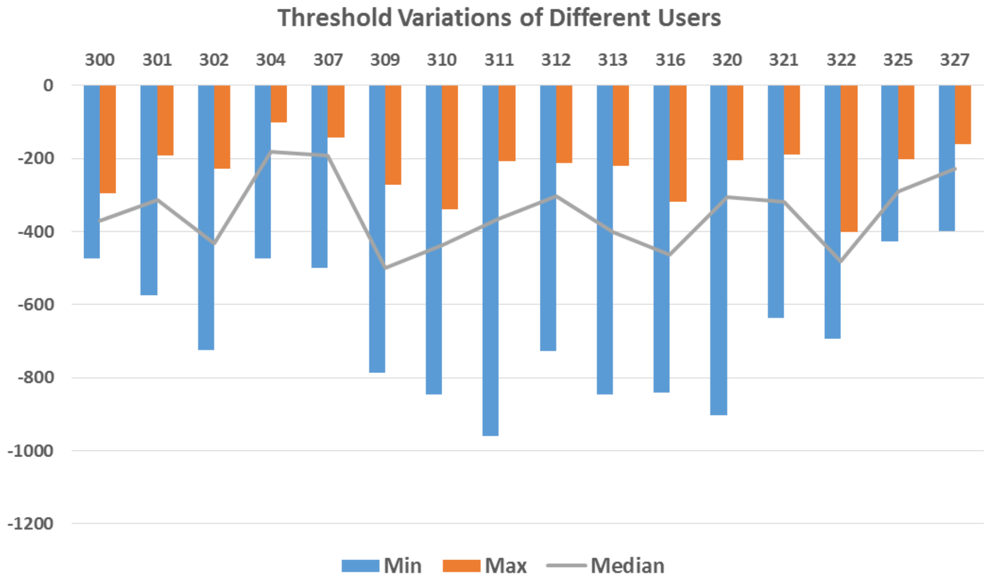


Figure 4.20: Distribution of the different thresholds: Minimum, Maximum and Median from all users

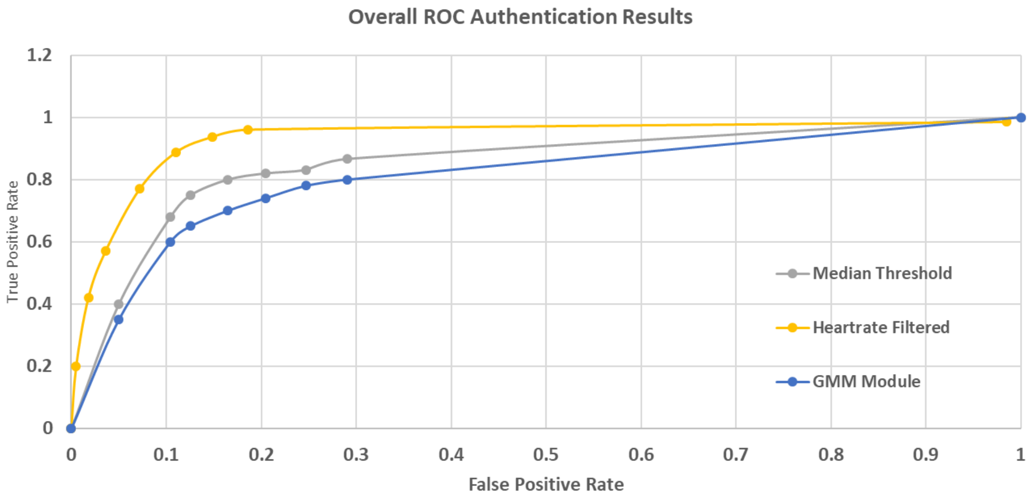


Figure 4.21: Relative Operating Characteristic of the identification process using median threshold, proposed architecture and GMM Module only

The results are tabulated and represented in ROC as shown in Figure 4.21. The results using user specific median threshold value achieved a 68% True Positive Rate with a corresponding 10% False Positive Negative Rate. Compared to using maximum loglikehood score, the False Positive Rate improved by double to achieve a False

Positive Rate of 10%

4.3.4 User Classification with User-Specific Heartrate and Thresholding Criteria

To further improve the classification, the use of individual heartrate in the identification process is proposed. Every individual has confined range of heartrate between resting and maximum duress. Hence, the individual loglikelihood score is constructed against individual heartrate. Figure 4.19 is replot into Figure 4.22 which illustrates the loglikelihood score against increasing heartrate. It is evident that the loglikelihood score decrease linearly with accelerating heartrate.

With the understanding of the behavior, it is proposed to further improve the classification process, two additional criteria can be imposed on the classification criteria. Prior to discriminate the individual based on highest loglikelihood score, it has to satisfy two criteria.

- Criteria 1: The heartrate from the sequence under test is within the range of the user.
- Criteria 2: The loglikelihood score from sequence under test must fall with the range of the user unique heartrate based loglikelihood score.

Hence, if the tested loglikelihood score is the highest and matches the two criteria, the result is deemed true positive. See Figure 4.23 for an illustration of the classification using heartrate and dynamic thresholding criteria

Through the use of heartrate in the classification, the identification improves by more than 35% to achieve a 0.89 True Positive Rate while maintaining a False Positive Rate of 0.11. Figure 4.24 shows the performance comparison between proposed architecture and fiducial based Linear Discriminant Analysis (LDA) with Nearest

Center as classifier. The LDA is commonly used in machine learning for biometrics verification and identification. To enable fair comparison with the LDA methodology, similar database from the proposed method is used. The performance of LDA achieved a TPR of approximately 0.78 and FPR of 0.25. In another comparison, GMM-HMM with user specific median score as criteria achieved 0.68 for TPR and 0.11 for FPR. This reinforces the hypothesis that the use of heartrate together with user specific thresholding criteria is crucial to achieve better identification accuracy under highly variated heartwave signals.

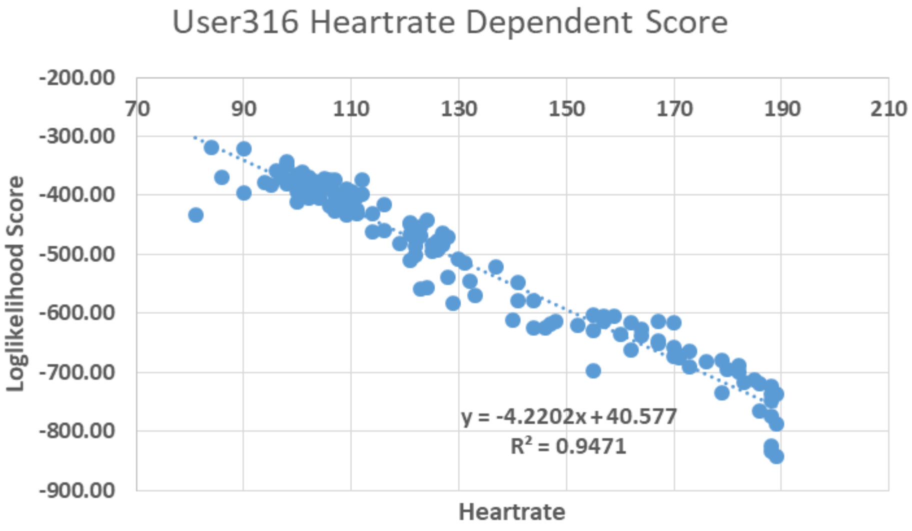


Figure 4.22: Decreasing loglikelihood score against increasing heartrate

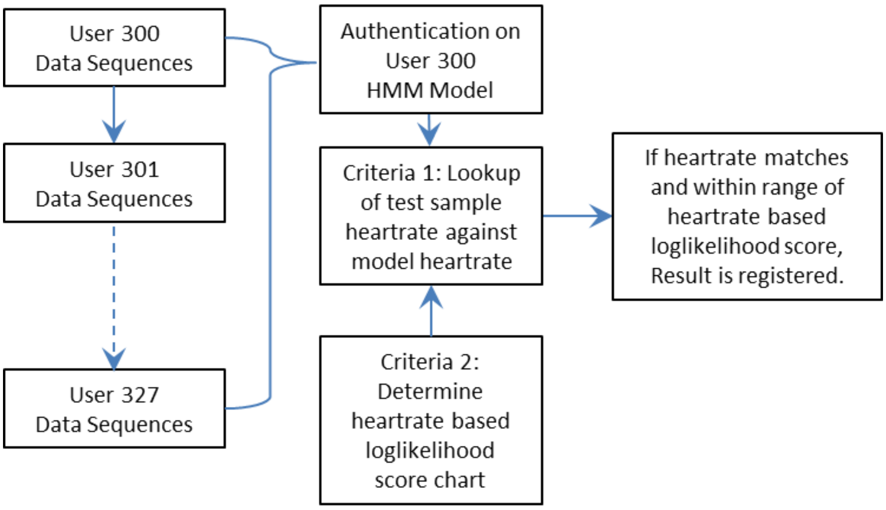


Figure 4.23: Schematic of classification using heartrate and dynamic thresholding criteria

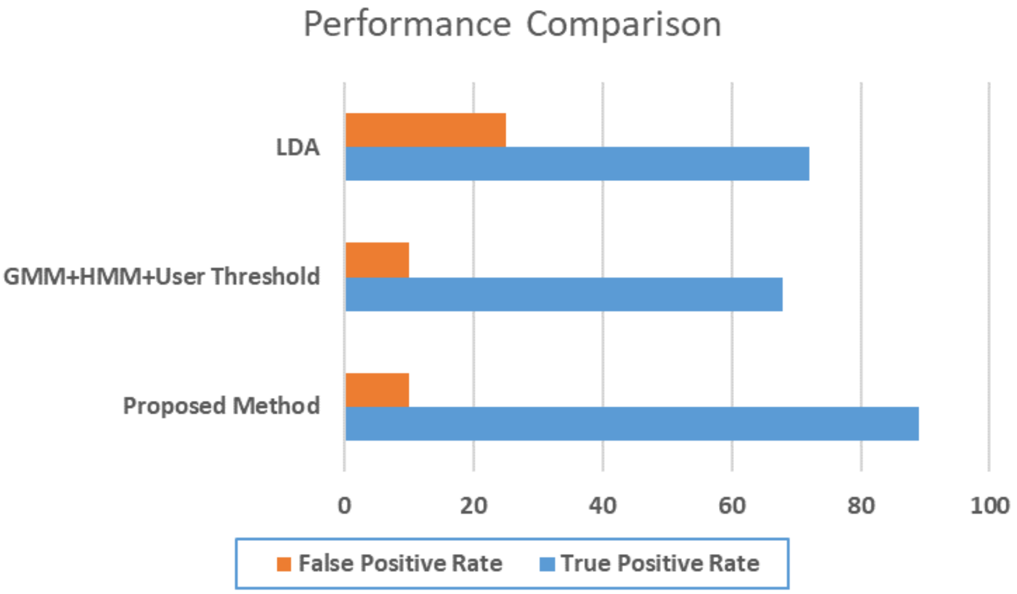


Figure 4.24: Performance of proposed architecture against other methods

4.3.5 Impact of Performance Between Resting and Intense Heartrate

The effect of heartwave morphological changes from heartrate variation (rest state to intense duress state) on the performance of the proposed architecture is investigated. Resting heartrates from all users are sorted and tested on the proposed architecture, and compared against LDA approach. Using resting heartrate dataset, the EER for the proposed architecture and LDA approach are relatively similar at 0.03 and 0.035 respectively. However, with the inclusion of heartwave data under intense heartrate, the EER for proposed architecture and LDA approach are 0.11 and 0.25 respectively. The performance of the proposed architecture has performed reasonably well at EER 0.11 on heartwave signal under full spectrum of heartrate variation. The result is shown in Figure 4.25.

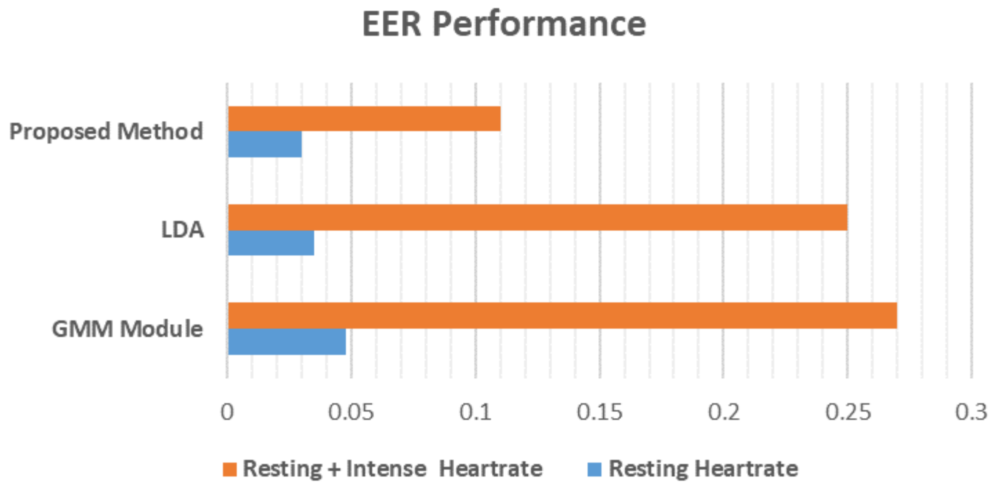


Figure 4.25: Comparison of proposed method against fiducial based LDA

4.4 Summary

The identification of individuals using heartwave has shown promising results. In the extraction process, it has proven that the use of heartrate adaptive parameters such as QT interval from Nomogram and PR interval have led to reliable extractions of heartwave features. In the preparation of data for classification, it is concluded that the use of BIC and MDL to limit the components do not significantly contribute to better classification results. In GMM-HMM classification testing, results have shown that the classification performs better at lower number of states than in higher number of states. The classification work to support identification achieved an ERR of 0.11.

It has been shown that using User-Specific Heartrate and Thresholding Criterial yielded a much desirable performance. Deliberately, identification performance is performed with just the GMM module using the same dataset. The EER based on GMM module alone is approximately 0.27. This underlines the importance of HMM module to achieve better identification performance. The work is limited by the availability of heartwave data where user has been subjected to physiological duress. However at the current development, the results have shown the feasibility of using heartwave signal as biometric mode with varying heartrate. This work has demonstrated that at varying heartrate, the heartwave signal exhibited unique characteristic features that can be used to discriminate individual.

CHAPTER 5

DEEP ENSEMBLE HEARTWAVE AUTHENTICATION

5.1 Introduction

Chapter 4 presented the statistical architecture of Gaussian Mixture Model–Hidden Markov Model (GMM-HMM) methodology where individual data are statistically clustered and modelled to support classification. In Chapter 5, the use of neural network based methodology to support heartwave based authentication is presented. Chapter 5 commences with the detailed explanation of the proposed architecture that consists of an ensemble-Deep Belief Network (DBN) framework, complemented by Multiview Spectral Embedding [106] connected to a single stacked DBN and classified using an efficient method of Extreme Learning Machine [122-124]. Equally, the performance of the proposed architecture is also compared against statistical methodology of GMM-HMM. In addition, this chapter will also present a second neural architecture known as Architecture 2 which exhibits similar classification performance to the proposed architecture.

5.1.1 Motivation

- To investigate the feasibility of proposed Deep Learning architecture to achieve a robust heartwave based biometric classification under extreme signal morphological variations.
- To investigate the uniqueness of anomaly heartwave signal which can be used as an authentication mode. This allows the possibility of heartwave based authentication solutions to be extended beyond healthy and normal heartwave signal.

- To investigate the minimum data requirement to support training of Deep Learning architecture to address the insufficiency of data for training.

5.2 Proposed Architecture of Deep Multi-View Heartwave Authentication

This chapter proposed a heartwave based authentication approach that is reliable and robustness to individual varying heartrate. Figure 5.1 illustrates the proposed architecture and can be broadly classified into 5 stages. Stage 1 focuses on the data preparation prior for input to the ensemble-DBN. Stage 2 focuses on the use of ensemble-DBN as opposed to using single DBN for feature extraction. Stage 3 focuses on the combination of outputs from ensemble-DBN into a single view and the adoption of Multi-View Spectral Embedding [106] method to form a single low dimensional embedding. Stage 4 focuses on the use of stacked DBN to support the eventual classification at Stage 5. The last stage focuses on the use of an efficient method of Extreme Learning Machine (ELM) for accurate and reliable classification.

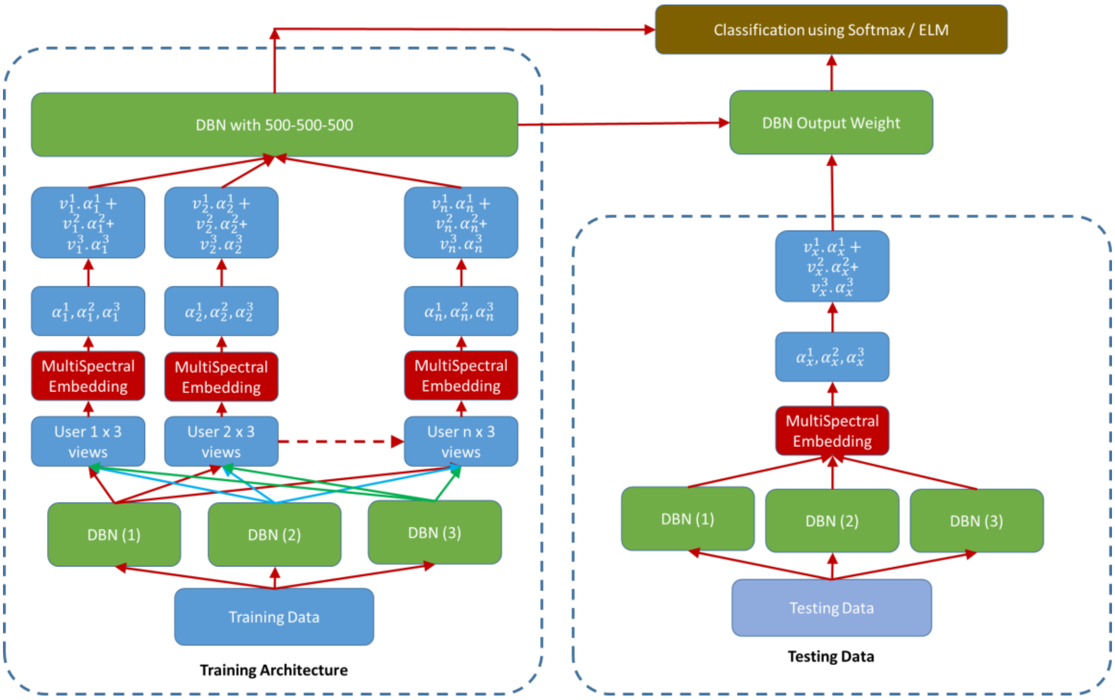


Figure 5.1: Illustration of the proposed deep multi-view architecture

5.2.1 Stage 1: Data preparation

In the data preparation, the heartwave of each individual is independently delineated and extracted. The delineation process uses Discrete Wavelet Transform (DWT) with a fourth order Daubechies wavelet (DB4) to achieve heartwave delineation. Chapter 3 has described in details the methodology to delineate each heartwave independently. To recap from Chapter 3, due to the presence of wide variations in heartrates, waves in particular to P-Wave and T-Wave experience significant morphological variations which are highly dependent on heartrate. Refer to Figure 5.2 for illustration. At elevated heartrate, the duration of T-Wave can shrink by approximately 40%. To ensure accurate delineation, the extraction of the heartwave signal incorporated heartrate related parameters of QT Interval and PR Interval to perform extraction of features related to P-Wave and T-Wave. The detail methodology of the heartwave delineation can be found in Chapter 3.

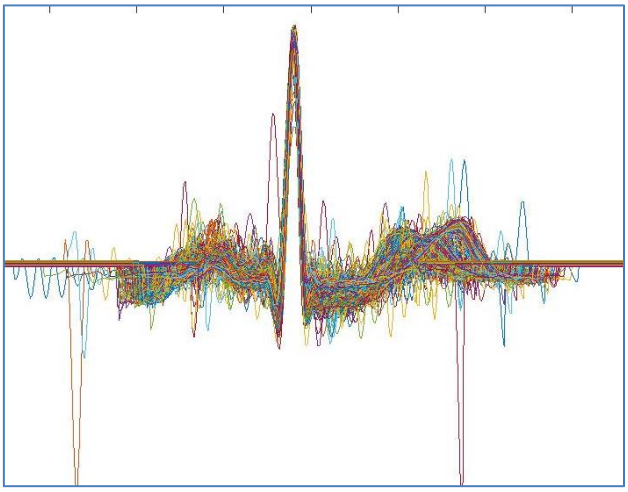


Figure 5.2: Superimpose of all extracted heartwaves of an individual around R-Peak. P-Wave and T-Wave show the morphological changes at different heartrate

For the investigation of the proposed architecture, heartwave signals from all 52 individuals consisting of 75,188 heartwaves are made available. Of the 52 individuals, there are 5 individuals with inverted T-Waves whose heartwaves have been successfully extracted. To further prepare the heartwaves for subsequent ensemble Deep Belief Network (DBN) training, each of the heartwave is aligned inside a fixed window length of 500 units. Using the location of the detected R-Peak, the R-Peak is aligned at the 240th unit. This provides a consistent data to the ensemble-DBN. See Figure 5.2 for an illustration of the alignment of heartwaves at about R-Peaks at 240th unit. In Deep Learning analysis, the data preparation differs from the data preparation for the statistical methodology. Recall in Chapter 4 on the statistical methodology, only the features of the heartwave signal are use and a total of 11 features are extracted. For the Deep Learning methodology, the extraction methodology presented in Chapter 3 is used to detect the heartwave extremities of P-Wave onset and T-Wave offset. This two features define the whole heartwave profile of each heartbeat. Upon detection of the two extremities, the signal data are extracted, store as a sample and concatenated into a data array. Included in the extraction is also the location of the R-Peak which will be used for alignment within a defined window.

5.2.2 Stage 2: Ensemble-Deep Belief Network (Ensemble-DBN)

Deep Belief Network (DBN) is an undirected probabilistic model that is constructed by multiple layers of Restricted Boltzmann Machine (RBM). RBM is a shallow stochastic neural network comprises of one layer of visible units and one layer of hidden units. Characteristically, each visible unit is connected to all hidden units and vice versa. Hence RBM is bipartite graph since no visible unit is connected to any other visible unit and is equally said for hidden layer. As a parameterized model of probability distribution, it is used to learn and identify important characteristics or features of an unlabeled target distribution based on the samples from the target distribution. Given a sample, the RBM learns and trains through the adjustment of its parameters to achieve a probability distribution that representatively fits the training data. RBM as a stochastic system is commonly used for pattern recognition and pattern classification.

Figure 5.3 illustrates a RBM layer. RBM, an undirected graphical model is also a bipartite undirected graph. It is contains *m* visible units and *n* hidden units. Each of the visible units capture the 1 units of the heartwave window and each hidden unit captures the dependency between units. The computation of RBM involves the use of energy model which is defined in 5.1 below.

$$E_{RBM}(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i h_j w_{ij} \tag{5.1}$$

From (5.1), $i \in \{1,2,\dots n\}$ and $j \in \{1,2,\dots m\}$ while *a* and *b* refer to the visible and hidden biases respectively. w_{ij} is the connecting weight between the visible and hidden units. The stated equation implies that given the parameters $\{a,b,w\}$, the energy between the visible layer and hidden layer can be computed.

Hence, the objective of the training RBM is to determine the parameters that best represent the probability distribution of the training data.

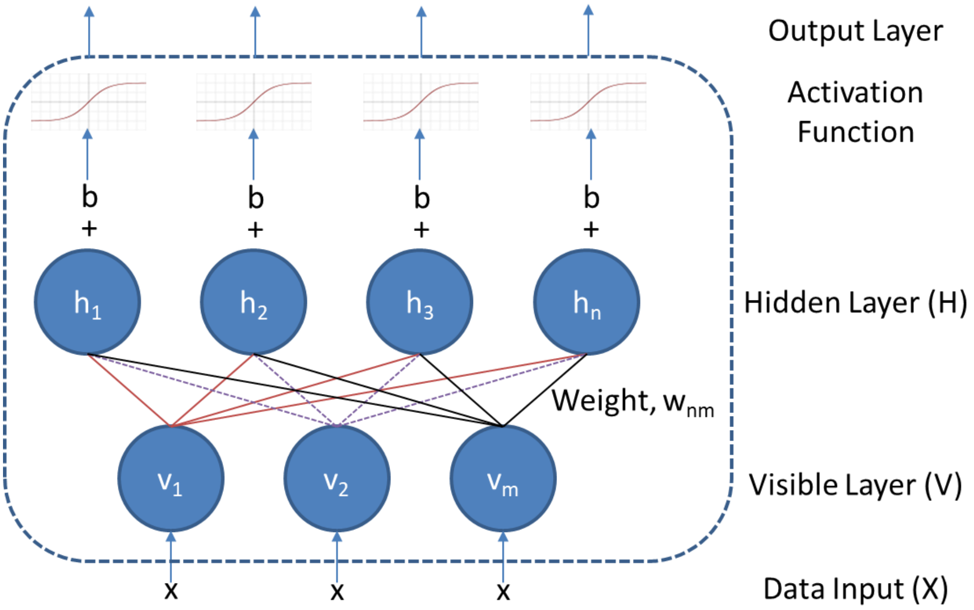


Figure 5.3: Illustrates a single layer RBM

Figure 5.3 also illustrates that only one layer is connected to one layer. Units within the same layer are not connected. This implies if the states of the hidden units and visible units are known, the visible units and hidden units are deem independent respectively. Through this understanding, the conditional distributions of $p(v|h)$ and $p(h|v)$ for each unit can be factorized as follows

$$p(v|h) = \prod_{i=1}^{m} p(v_i|h) \tag{5.2}$$

$$p(h|v) = \prod_{j=1}^{n} p\left(h_j|v\right) \tag{5.3}$$

Based on Gibbs distribution and energy model, the joint distribution of the visible and hidden unit can be described as follow:

$$p(v,h) = \frac{1}{Z}\exp\left(-E_{RBM}(v,h)\right) \tag{5.4}$$

$$Z = \sum_{v,h} \exp(-E_{RBM}(v,h)) \tag{5.5}$$

where *Z* is known as the partition function and is the summation of all possible pairs of visible and hidden vectors. The probability that the network assigns to a visible vector is given by summing over all possible hidden vector:

$$p(v) = \sum_{h} p(v,h) = \frac{1}{Z}\sum_{h} \exp\left(-E_{RBM}(v,h)\right) \tag{5.6}$$

Let the parameters of the RBM layer be represented by $\theta = \{w, a, b\}$. Expression (5.6), based on given parameters, it can be expressed as follow,

$$p(v|\theta) = \sum_{h} p(v,h) = \frac{1}{Z}\sum_{h} \exp\left(-E_{RBM}(v,h)\right) \tag{5.7}$$

Hence, to estimate the parameters of the RBM model, maximum likelihood can be applied to expression (5.7).

$$L(\theta|v) = (p(v|\theta) = \sum_h p(v,h) = \frac{1}{Z}\sum_h \exp\big(-E_{RBM}(v,h)\big) \qquad (5.8)$$

Applying log to expression (5.8),

$$logL(\theta|v) = logp(v|\theta) = log\frac{1}{Z}\sum_h \exp\big(-E_{RBM}(v,h)\big)$$

$$= log\sum_h \exp\big(-E_{RBM}(v,h)\big) - log\sum_{v,h} \exp(-E_{RBM}(v,h)) \qquad (5.9)$$

Applying a partial derivative with respect to the parameter, (5.9) becomes

$$\frac{\partial logL(\theta|v)}{\partial\theta} = \frac{\partial}{\partial\theta}\left(log\sum_h exp\big(-E_{RBM}(v,h)\big)\right)$$

$$-\frac{\partial}{\partial\theta}\left(log\sum_{v,h} exp\big(-E_{RBM}(v,h)\big)\right)$$

$$= \frac{1}{\sum_h exp\big(-E_{RBM}(v,h)\big)}\sum_h exp\big(-E_{RBM}(v,h)\big)\left(-\frac{\partial E_{RBM}(v,h)}{\partial\theta}\right)$$

$$-\frac{1}{\sum_{v,h} exp\big(-E_{RBM}(v,h)\big)}\sum_{v,h} exp\big(-E_{RBM}(v,h)\big)\left(-\frac{\partial E_{RBM}(v,h)}{\partial\theta}\right)$$

$$= -\sum_h \frac{exp\big(-E_{RBM}(v,h)\big)}{\sum_h exp\big(-E_{RBM}(v,h)\big)}\frac{\partial E_{RBM}(v,h)}{\partial\theta}$$

$$+\sum_{v,h} \frac{exp\big(-E_{RBM}(v,h)\big)}{\sum_{v,h} exp\big(-E_{RBM}(v,h)\big)}\left(\frac{\partial E_{RBM}(v,h)}{\partial\theta}\right)$$

$$= -\sum_h p(h|v)\left(\frac{\partial E_{RBM}(v,h)}{\partial\theta}\right) + \sum_{v,h} p(v,h)\left(\frac{\partial E_{RBM}(v,h)}{\partial\theta}\right) \qquad (5.10)$$

The condition probability term $p(h|v)$ can be solved by the following

$$p(h|v) = \frac{p(v,h)}{p(v)} = \frac{\frac{1}{Z}\exp\big(-E_{RBM}(v,h)\big)}{\frac{1}{Z}\sum_h \exp(-E_{RBM}(v,h))} = \frac{\exp\big(-E_{RBM}(v,h)\big)}{\sum_h \exp\big(-E_{RBM}(v,h)\big)} \tag{5.11}$$

Expression (5.10) comprises of two terms. The first term represents the probability value of the energy from the conditional probability of the hidden units based on the training samples. The second term refers to the joint distribution between the hidden and visible units and it can be expressed as $\sum_v p(v) \sum_{v,h} p(h|v) \left(\frac{\partial E_{RBM}(v,h)}{\partial \theta}\right)$. This expression however is computationally extensive and complex.

Given the expression of (5.11), it is now possible to compute the each of the parameters of the RBM model $\{w, a, b\}$. Hence, individual parameters of the model can be computed by imposed a derivative with respective parameter components.

$$\begin{aligned}\frac{\partial logL(\theta|v)}{\partial w_{ij}} &= -\sum_h p(h|v)\left(\frac{\partial E_{RBM}(v,h)}{\partial w_{ij}}\right) + \sum_{v,h} p(v,h)\left(\frac{\partial E_{RBM}(v,h)}{\partial w_{ij}}\right)\\ &= \sum_h p(h|v) h_i v_j - \sum_v p(v) \sum_h p(h|v)\, h_i v_j\\ &= p(h_i = 1|v) v_j - \textstyle\sum_v p(v)\, p(h_i = 1|v) v_j \end{aligned} \tag{5.12}$$

where from (5.1)

$$\begin{aligned}\frac{\partial E_{RBM}(v,h)}{\partial w_{ij}} &= \frac{\partial}{\partial w_{ij}}\left(-\sum\nolimits_i a_i v_i - \sum\nolimits_j b_j h_j - \sum\nolimits_{ij} v_i h_j w_{ij}\right)\\ &= -h_i v_j \end{aligned} \tag{5.13}$$

And as the hidden state is a binary output of sigmoid function, the state of the hidden unit is either ON or OFF with $h_i \in \{0,1\}$,

$$\sum_h p(h|v)h_i = p(h_i = 1|v) \tag{5.14}$$

Similarly, the parameters of a and b for the i^{th} and j^{th} variables can be computed using the respective derivatives.

$$\frac{\partial logL(\theta|v)}{\partial a_j} = -\sum_h p(h|v)\left(\frac{\partial E_{RBM}(v,h)}{\partial a_j}\right) + \sum_{v,h} p(v,h)\left(\frac{\partial E_{RBM}(v,h)}{\partial a_j}\right)$$

$$= \sum_h p(h|v)v_j - \sum_v p(v)\sum_h p(h|v)\, v_j$$

$$= v_j - \sum_v p(v)\, v_j \tag{5.15}$$

$$\frac{\partial log\mathcal{L}(\theta|v)}{\partial b_i} = -\sum_h p(h|v)\left(\frac{\partial E(v,h)}{\partial b_i}\right) + \sum_{v,h} p(v,h)\left(\frac{\partial E(v,h)}{\partial b}\right)$$

$$= \sum_h p(h|v)h_i - \sum_v p(v)\sum_h p(h|v)\, h_i$$

$$= p(h_i = 1|v) - \sum_v p(v)\, p(h_i = 1|v) \tag{5.16}$$

The states of the single variable being turn ON depend on conditional probability of a single variable with sigmoid activation function $sigm(x) = \frac{1}{1+exp(-x)}$.

$$p(h_i = 1|v) = sigm\left(\sum_{j=1}^{m} w_{ij}v_j + b_i\right) \tag{5.17}$$

$$p\left(v_j = 1\middle|h\right) = sigm\left(\sum_{i=1}^{n} w_{ij}h_i + a_j\right) \tag{5.18}$$

(5.17) and (5.18) can be derived as follows. Let v_{-k} represents the states of units that are visible except for the k-th one and hence can be defined as

$$\alpha_k(h) = -\sum_{i=1}^{n} w_{ik}h_i - b_k$$

$$\beta(v_{-k},h) = -\sum_{i=1}^{n}\sum_{j=1,j\neq k}^{m} w_{ij}h_i v_j - \sum_{j=1,j\neq k}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i$$

Therefore $E_{RBM}(v,h) = \beta(v_{-k},h) + v_k\alpha_k(h)$, with $v_k\alpha_k(h)$ represents the terms involving v_k.

$$\begin{aligned}
p(v_k = 1|h) = p(v_k = 1|v_{-k},h) &= \frac{p(v_k = 1, v_{-k}, h)}{p(v_{-l}, h)} \\
&= \frac{\mathrm{e}^{-E_{RBM}(v_k=1,v_{-k},h)}}{\mathrm{e}^{-E_{RBM}(v_k=1,v_{-k},h)} + \mathrm{e}^{-E_{RBM}(v_k=0,v_{-k},h)}} \\
&= \frac{\mathrm{e}^{-\beta(v_{-k},h)-1\cdot\alpha_k(h)}}{\mathrm{e}^{-\beta(v_{-k},h)-1\cdot\alpha_k(h)} + \mathrm{e}^{-\beta(v_{-k},h)-0\cdot\alpha_k(h)}} \\
&= \frac{\mathrm{e}^{-\beta(v_{-k},h)} \cdot e^{-\alpha_k(h)}}{\mathrm{e}^{-\beta(v_{-k},h)} \cdot \mathrm{e}^{-\alpha_k(h)} + \mathrm{e}^{-\beta(v_{-k},h)}} \\
&= \frac{\mathrm{e}^{-\beta(v_{-k},h)} \cdot e^{-\alpha_k(h)}}{\mathrm{e}^{-\beta(v_{-k},h)} \cdot \mathrm{e}^{-\alpha_k(h)+1}} \\
&= \frac{e^{-\alpha_k(h)}}{\mathrm{e}^{-\alpha_k(h)+1}} \\
&= \frac{1}{1 + e^{\alpha_k(h)}} \\
&= sigm\big(-\alpha_k(h)\big) \\
&= sigm\left(\sum_{i=1}^{n} w_{ik}h_i + b_k\right)
\end{aligned}$$

From (5.12), (5.15) and (5.16), of the two terms of the loglikelihood gradient, the first term on the left can be computed whereas the second term will least to complexity in the summing of visible units. An effective method to overcome this complexity is the adopting of approximation via Gibbs sampling. To enable Gibbs sampling to be achieved, it requires repetitive iterations to ensure convergence. An alternative is proposed through the use of Contrastive Divergence to determine the approximate the loglikelihood. Contrastive Divergence is always accompanied by k-step which denotes the number of steps for approximation computation. Contrastive Divergence is also commonly denotes as CD. The Gibbs chain starts with a training data $v^{(0)}$ of the training dataset and produces the sample $v^{(k)}$ after k step. Each step represented by t involves sampling $h^{(t)}$ from $p(h|v^{(t)})$ followed by sampling $v^{(t+1)}$ from $p(v|h^{(t)})$. The gradient (5.10) with respect to θ of the log-likelihood for the specific sample $v^{(0)}$ is represented by

$$CD_k\left(\theta, v^{(0)}\right) = -\sum_h p\left(h\middle|v^{(0)}\right)\frac{\partial E_{RBM}\left(v^{(0)},h\right)}{\partial\theta} + \sum_h p\left(h\middle|v^{(k)}\right)\frac{\partial E_{RBM}\left(v^{(k)},h\right)}{\partial\theta} \tag{5.19 }$$

Hence each of the parameters can be obtained by approximating the expectations over $p(v)$

$$\begin{aligned} CD_1\left(\mathrm{w}_{ij}, v^{(0)}\right) &= -\sum_h p\left(h\middle|v^{(0)}\right)\frac{\partial E_{RBM}\left(v^{(0)},h\right)}{\partial\theta} + \sum_h p\left(h\middle|v^{(1)}\right)\frac{\partial E_{RBM}\left(v^{(1)},h\right)}{\partial\theta} \\ &= p\left(h_i = 1\middle|v^{(0)}\right)v_j^{(0)} - p\left(h_i = 1\middle|v^{(1)}\right)v_j^{(1)} \end{aligned} \tag{5.20}$$

$$\begin{aligned} CD_1\left(\mathrm{a}_j, v^{(0)}\right) &= -\sum_h p\left(h\middle|v^{(0)}\right)\frac{\partial E_{RBM}\left(v^{(0)},h\right)}{\partial\theta} + \sum_h p\left(h\middle|v^{(1)}\right)\frac{\partial E_{RBM}\left(v^{(1)},h\right)}{\partial\theta} \\ &= v_j^{(0)} - v_j^{(1)} \end{aligned} \tag{5.21}$$

$$\begin{aligned} CD_1\left(\mathrm{b}_i, v^{(0)}\right) &= -\sum_h p\left(h\middle|v^{(0)}\right)\frac{\partial E_{RBM}\left(v^{(0)},h\right)}{\partial\theta} + \sum_h p\left(h\middle|v^{(1)}\right)\frac{\partial E_{RBM}\left(v^{(1)},h\right)}{\partial\theta} \\ &= p\left(h_i = 1\middle|v^{(0)}\right) - p\left(h_i = 1\middle|v^{(1)}\right) \end{aligned} \tag{5.22}$$

As the heartwave signal contains continuous data, it is necessary to prep the signals where the input to the visible layer are linear units with independent Gaussian noise.

Therefore the energy function of (5.1) becomes

$$E_{RBM}(v,h) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}h_i\frac{v_j}{\sigma_j} - \sum_{j=1}^{m}\frac{\left(v_j - a_j\right)^2}{2\sigma_j^2} - \sum_{i=1}^{n} c_i b_i \tag{5.23}$$

This RBM is termed as Gaussian Bernoulli RBM (GBRBM) and the hidden units remains as binary. Since the visible units contains linear units with Gaussian noise, the updated rules for visible units becomes

$$p\left(v_j \middle| h\right) = \mathcal{N}\left(v_j \middle| \mu_j, \sigma_j^2\right) \tag{5.24}$$

where $\mu_j = \sum_{i=1}^{n} w_{ij}h_i + b_j$ and $\mathcal{N}$ is normal distribution.

This concludes the unsupervised pre-training for a RBM. In a single DBN, three layers of RBM are stack onto each other. To achieve a supervised trained DBN, an output layer consisting of 52 nodes is stacked above the third layer. Once the initial set of parameters for the first layer has been determined by the pre-training, another layer of RBM is stacked and the activation probabilities of the first RBM are used as input data for the second RBM. Similarly, pre-training continues until the second set of initial parameters for the second layer is completed. Pre-training continues till the third layer of RBM. See Figure 5.4 for illustration of a DBN. Upon completion of pre-training for the three layers of RBM, the pre-determined initial weights in the three layers are subjected to supervise training using backpropagation with fine tuning with the output layer using commonly adopted softmax method. The pre-training on the stack RBM is an unsupervised model, which characterizes the input data distribution using hidden variables and there is no label information provided.

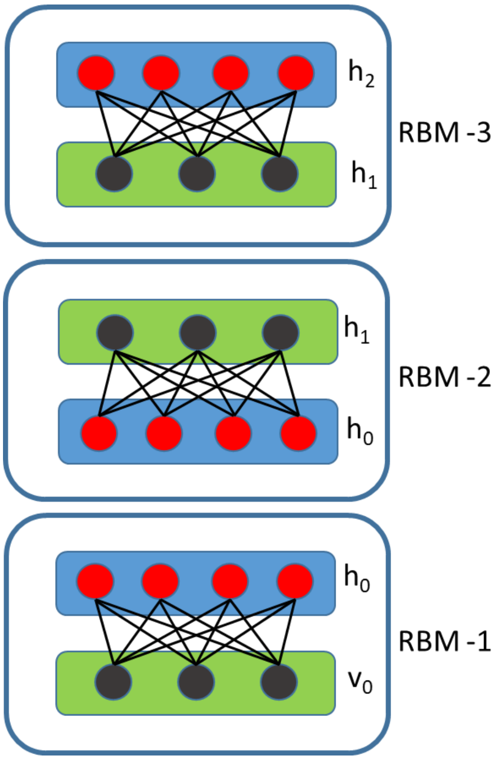


Figure 5.4: Illustrates a pre-train DBN with 3 layers of RBM

Although the parameters of the 3-layer RBM has been obtained using layer-wise pre-training method. The parameters are not optimal for classification. Therefore, supervised fine-tuning is required to optimize the parameters to improve the performance of the network. To support the objective of classification, the fine-tuning is performed by adding a layer of variables that represent the desired labels provided in the training data. Back-propagation algorithm is implemented to fine tune the parameters using the approach of feed-forward method in neural network. See Figure 5.5 for illustration of a supervised fine-tuning.

The classification layer uses Softmax as the activation functions. As the training target for classes of users is value either 0 or 1, softmax function is adopted. The output layer as softmax function is a generalization of logistic function, whose output represents a categorical distribution.

$$p(y = j|x) = \frac{e^{x^{\mathrm{T}} w_j}}{\sum_{k=1}^{K} e^{x^{\mathrm{T}} w_k}} \tag{5.25}$$

where $p(y = j|x)$ denotes the predicted probability for the jth class given a sample vector x and a weighting vector w.

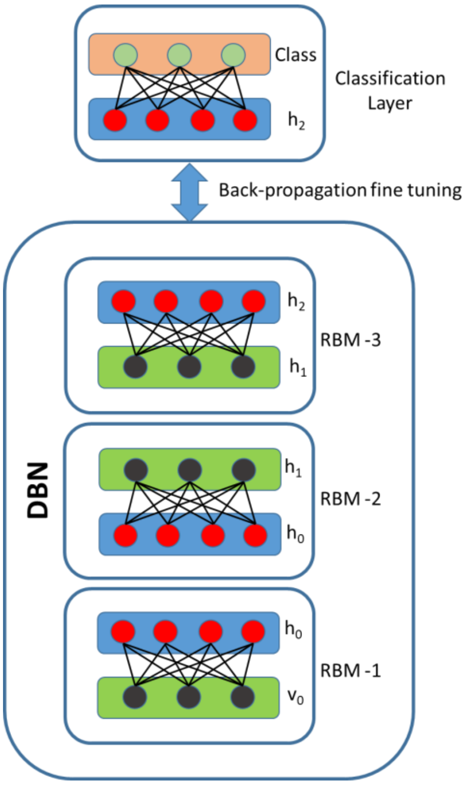


Figure 5.5: DBN with classification layer for supervised training

In the proposed method, 3 separate DBNs are adopted to form an ensemble-DBN. The continuous data of heartwave signal in particular to records acquired under physical duress contains multiple high frequency components and random spikes possibly contributed by contacts of the signal electrodes. To address signal noises and the morphological changes due to physical duress with the objective to ensure maximum reliability for feature extraction, 3 DBNs with different parameters are used to maximize features extraction. The ensemble-DBN consist of the following configurations: DBN-1 with 500-100-500-52, DBN-2 with 500-500-500-52 and DBN-3 with 500-1000-500-52 where the last layer is the output layer for softmax based classification. The configuration of the ensemble-DBN including the number of DBNs are determined by hyper-parameters optimization using parameters of the number of hidden nodes, number of hidden layers and number of DBN based on the results of False Acceptance Rate and False Rejection Rate The inputs to the three DBNs are identical and the outputs from the penultimate layers of the multiple-DBNs are subsequently input into Stage 3 for low dimensional embedding using multi-view spectral methodology. See Figure 5.6 for illustration.

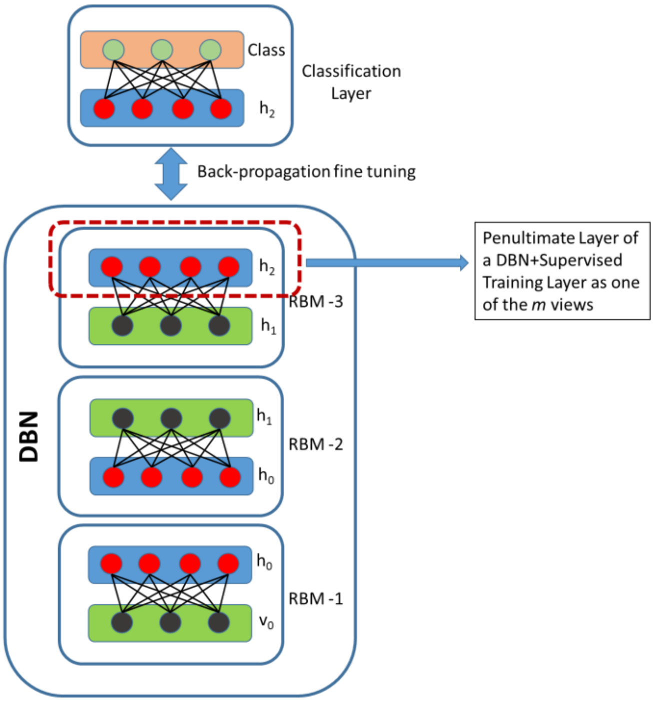


Figure 5.6: Penultimate layer of DBN as input to MSE

5.2.3 Stage 3: Multi-View Spectral Embedding

Outputs of the penultimate layers of the ensemble-DBN contain the extracted features under different DBN configurations. Conversely, the outputs also contain features that are irrelevant and contribute to the misclassification rate. Notably, in low training data, the error in misclassification increases exponentially. The use of multi-view spectral embedding (MSE) is an effective approach to combine representations of multi-DBN through the identification of the complementary property of each view and embed to form a singular view. See Figure 5.7 for illustration of the MSE methodology.

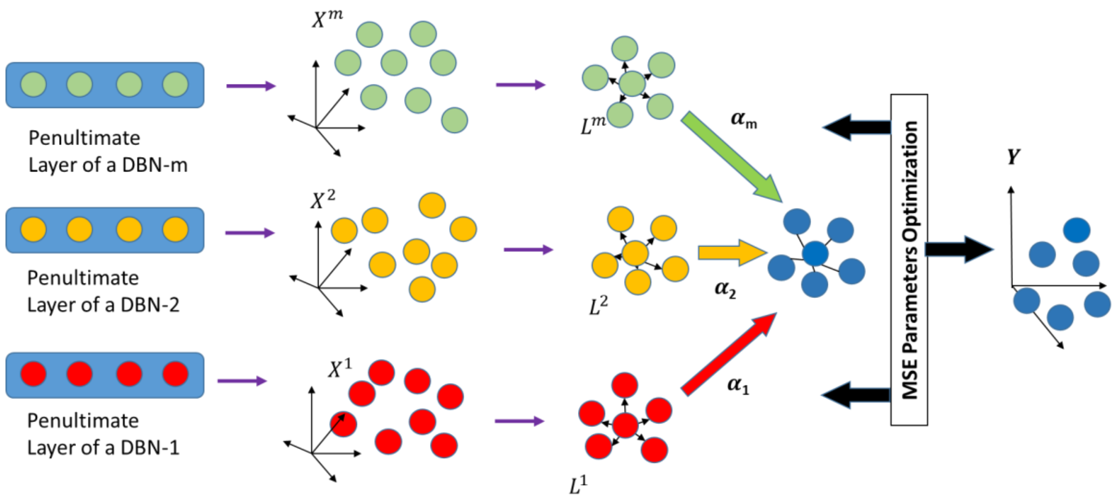


Figure 5.7: Schematic of MSE methodology

Prior to input to MSE, the data is sort according to individual users. Hence, the data are sorted into 52-user specific clusters comprising of 3 views each. Each of 52 clusters is individually input to MSE to determine the contribution factor of each view per user specific cluster $\alpha = [\alpha_1, \dots, \alpha_m]$ where m refers to the number of DBNs in the ensemble-DBN.

Let X be the representation from each of the DBN input into the MSE module and is given by $X = \{X^m \in \mathbb{R}^{m_i \times n}\}_{i=1}^{m}$ where n is the number of objects or samples. X^m is the feature matrix from each DBN. For each of the feature matrix,

$$X^i = \left[x_{1,}^i \dots . x_n^i\right] \in \mathbb{R}^{m_i \times n} \tag{5.26}$$

In each of the view, a patch is determined by considering an arbitrary point with its k-related samples using nearest neighbor. Hence, consider an arbitrary point x_j^i

together with its k related points, the patch can be defined as $X_j^i = \left[x_{j,}^i x_{j1,}^i \dots . x_{jk}^i\right] \in \mathbb{R}^{m_i \times (k+1)}$. For each of the determined patch, a mapping is performed and express as $f_j^i : X_j^i \rightarrow Y_j^i$ where $Y_j^i = \left[y_{j,}^i y_{j1,}^i \dots . y_{jk}^i\right] \in \mathbb{R}^{d \times (k+1)}$ from each view, the part mapping is defined as

$$\underset{Y_j^i}{\operatorname{argmin}} \sum_{l=1}^{k} \left\| y_j^i - y_{jl}^i \right\|^2 \left(\omega_j^i\right)_l \tag{5.27}$$

where ω_j^i is a dimensional column vector weighted limited by k. The vector weight is represented by $\left(\omega_j^i\right)_l = exp\left(- \left\|x_j^i - x_{jk}^i\right\|^2 / \gamma\right)$. The closeness or the compactness of the neighborhoods is defined by γ. Hence, the optimization of mapping can be expressed to

$$\underset{Y_j^i}{\operatorname{argmin}}\, tr\left(\begin{bmatrix} \left(y_j^i - y_{jl}^i\right)^T \\ \dots \\ \left(y_j^i - y_{jk}^i\right)^T \end{bmatrix} \times \left[y_j^i - y_{jl}^i, \dots, y_j^i - y_{jk}^i\right] diag\left(\omega_j^i\right)\right)$$

$$= \underset{Y_j^i}{\operatorname{argmin}}\, tr\left(Y_j^i L_j^i \left(Y_j^i\right)^T\right) \tag{5.28}$$

where $tr(\cdot)$ is the trace operator, $L_j^i = \begin{bmatrix} \sum_{i=1}^{k} \left(\omega_j^i\right)_l & -\left(\omega_j^i\right)^T \\ -\omega_j^i & diag\left(\omega_j^i\right) \end{bmatrix} \in \mathbb{R}^{(k+1) \times (k+1)}$ that embeds the complimentary property of a view into a low-dimensional embedding Y_j^i. As MSE module is managing multiple views, there is a need to assign a weight factor to each of the represented views. This is to ensure that only complimentary and homogenous property of multiple views contributes to the embedment. To extract the complimentary property, nonnegative weights $\alpha = [\alpha_1, \dots, \alpha_m]$ are incorporated to each of the optimized mapping of each view. The significance of the complement property of a view is directly proportional to the value of the non-negative weight factor α_i. Hence with the inclusion of weight to the represented m-th learned views, the summation of the jth part from the all views can be expressed as

$$\underset{Y=\left\{Y_j^i\right\}_{i=1}^{m},\alpha}{\operatorname{argmin}} \sum_{i=1}^{m} \alpha_i tr\left(Y_j^i L_j^i \left(Y_j^i\right)^T\right) \tag{5.29}$$

To ensure all patches are reference to a specific global origin, Y_j^i is mapped using the coordinates of $Y_j^i = [y_{j,}^i y_{j1,}^i \ldots . y_{jk}^i]$ which is referred globally to $Y = [y_1, y_2, y_3, \ldots, y_n]$. $Y_j^i = YS_j^i$, where $S_j^i \in \mathbb{R}^{n\times(k+1)}$ is the selection matrix to translate the relationship of parts from its original space into a patch. Therefore, (5.29) can be rewritten as

$$\underset{Y_j,\alpha}{\operatorname{argmin}} \sum_{i=1}^{m} \alpha_i tr\left(YS_j^i L_j^i \left(S_j^i\right)^T (Y)^T\right) \tag{5.30}$$

By summing all the optimizations of the part, the globally aligned parts should be expressed as

$$\begin{cases} \underset{Y,\alpha}{\operatorname{argmin}} \sum_{i=1}^{m} \alpha_i^{\varepsilon} tr\left(YL^iY^T\right) \\ \quad s.t. YY^T = I, \sum_{i=1}^{m} \alpha_i^{\varepsilon} = 1, \alpha_i \geq 0 \end{cases} \tag{5.31}$$

where $L^i \in \mathbb{R}^{n\times n}$ is the alignment matrix for the mth learned representations, and it is expressed as $L^i = \sum_{j=1}^{n} S_j^i L_j^i \left(S_j^i\right)^T$. For the Y, it can be determined by constraining YY to $YY = I$. ε controls the relationships between views where $\varepsilon \geq 1$. Normalizing L^i, the Laplacian Graph L_{sys} is symmetric and is expressed as follows:

$$L_{sys} = D^{i^{-\frac{1}{2}}} L^i D^{i^{-\frac{1}{2}}} = I - D^{i^{-\frac{1}{2}}} Q^i D^{i^{-\frac{1}{2}}} \tag{5.32}$$

where $Q^i \in \mathbb{R}^{n\times n}$ and $\left[Q^i\right]_{lj} = exp\left(-\left\|x_n^i - x_j^i\right\|^2/\gamma\right)$ wherer x_n^i is part of the group in k-nearest neighbors of x_j^i and $\left[Q^i\right]_{lj} = 0$ if it is not. D^i is diagonal matrix of degree $\left[D^i\right]_{jj} = \sum_l\left[Q^i\right]_{lj}$. To solve (5.30), the solution can be obtained by using Expectation Maximization (EM). The process is iterative through the optimizations of R and α.

Maximization: M-Step: Update α

λ, Lagrange multiplier is incorporated and constrain $\sum_{i=1}^{m}\alpha_i^\varepsilon$ to unity which can be expressed as shown:

$$L(\alpha,\lambda) = \sum_{i=1}^{m}\alpha_i^\varepsilon tr\left(YL_{sys}Y^T\right) - \lambda\left(\sum_{i=1}^{m}\alpha_i - 1\right) \qquad (5.33)$$

Imposing derivative on $L(\alpha,\lambda)$ with respect to α_i and λ to zero, α_i can be determined via (5.34) by proposing a fixed *Y*.

$$\alpha_i = \frac{\left(1/tr\left(YL_{sys}Y^T\right)\right)^{1/(\varepsilon-1)}}{\sum_{i=1}^{m}\left(1/tr\left(YL_{sys}Y^T\right)\right)^{1/(\varepsilon-1)}} \qquad (5.34)$$

Expectation: E-Step: Update Y.

To solve expression (5.31) optimization constraints, which are

$$\min_Y(YLY^T) \quad s.t. YY^T = I \qquad (5.35)$$

where $L = \sum_{i=1}^{m}\alpha_i^\varepsilon L_{sys}$. Using the theorem of KyFan, the optimized solution for expression (5.31) can be resolved by imposing a fixed α. The optimization of *Y* shall be represented by eigenvectors extracted from smallest eigenvalues equivalent the final embedded dimension from matrix L.

With the optimized α_i obtained, it is multiply to the penultimate layers of the respective views and summed. This ensures that for each user, only significant complementary property is amplified. This process is applies to all users.

5.2.4 Stage 4: Ensemble-DBN with Stacked-DBN

At this stage, a second DBN is stacked above the MSE module for extraction of higher order features. The input to the stacked DBN can be represented as follow

$$\Psi_m = f_{MSE}\big(D(\sigma_i) + \cdots + \mathrm{D}(\sigma_m)\big) \tag{5.36}$$

where i refers to the number of DBN in the ensemble-DBN and σ_i refers to the multiplied factor α_i to the penultimate layer of respective learned representation. The lower features from the ensemble-DBN are inputs to the stacked DBN which serve to identity higher order features. More importantly, views with higher significant contribution are given higher weightage relative the associated views. With the contribution of weightage from MSE, the view with significant domain features will provide greater learning probability by the DBN. This is in contrast to linear combination where the inputs from the three views are averaged which leads to significant domain features being suppressed and lesser significant domain features being elevated. The DBN used in this stack is configured with the following configuration of 500-500-500-52 where the last layer is the classification layer.

5.2.5 Stage 5: Classification with Extreme Learning Machine

From the penultimate layer of the stacked DBN, Extreme Learning Machine (ELM) is used as a classifier. See Figure 5.8 for illustration. The classification approach is not limited to ELM, approach like Softmax is applicable. The advantage of ELM over Softmax is the significantly lower computation speed required to train and test the samples. Another difference between ELM and Softmax lies on the training process. Softmax is an iterative process to fine tune the weights in the classification layer. ELM however is an analytical process that relies on universal approximation. ELM is essentially a single hidden-layer feedforward neural network (SLFN). ELM accomplishes through the assignment of random weights and biases to the hidden nodes and subsequently uses matrix computations to determine the output weights.

Given N as inputs to ELM where $\{x_{i,}t_i\}_{i=1}^{N}$. ELM model with K hidden nodes can be written as (5.37) and as shown in Figure 5.9

$$t_i = \sum_{\kappa=1}^{K} g(x_i \cdot u_i + v_\kappa)\,\beta_\kappa, \quad i = 1, \dots, N$$

$$= \textstyle\sum_{\kappa=1}^{K} g(x_i, u_\kappa, v_\kappa)\,\beta_\kappa \tag{5.37}$$

where t_i is the output, x_i denote the input vector, u_κ and v_κ are the parameters of the activation function of the κth hidden node, $g(x_\phi, u_\kappa, v_\kappa)$ is the output of the κth hidden node with respect to the κth input. β_κ is the output weight of the κth hidden node. The expression of (5.37) can be written as

$$\boldsymbol{T} = \boldsymbol{g\beta} \tag{5.38}$$

where $\boldsymbol{T} = [t_1, \dots, t_i]^T$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^{\boldsymbol{T}}$, and the hidden output matrix

$$\boldsymbol{g} = \begin{bmatrix} g(x_1, u_1, v_1) & \cdots & g(x_1, u_K, v_K) \\ \vdots & \ddots & \vdots \\ g(x_N, u_1, v_1) & \cdots & g(x_N, u_K, v_K) \end{bmatrix}_{N \times K} \tag{5.39}$$

An ELM learns the parameters in two stages: 1) random feature mapping and 2) linear parameter solving. In the first stage, the input data is projected into a feature space with randomly initialized parameters using the activation function$s(\cdot)$. It has been proven that the randomly initialized parameters are able to approximate any continual function [123, 124]. During the training of ELM, only two parameters are to be established. They are the number of hidden neurons and type of activation function. By default, typically the sigmoid function is selected as the activation function. Therefore, the only parameter that needs to be determined is the output weight $\boldsymbol{\beta}$, which can be estimated by

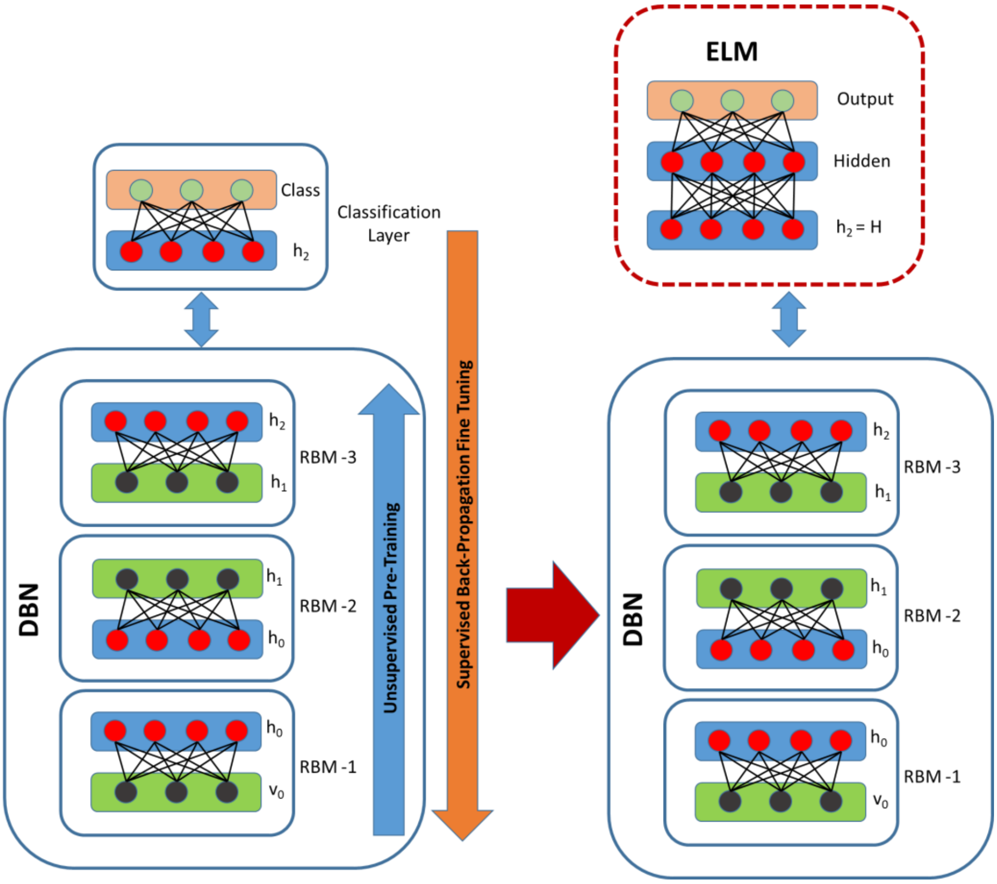


Figure 5.8: Adoption of ELM after supervised training

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{g}^{\dagger}\boldsymbol{T} \tag{5.40}$$

where $\boldsymbol{g}^{\dagger}$ is the Moore-Penrose generalized inverse. The use of ELM offers the following benefits that have been well documented by [122, 125, 126]. Advantages include local minimal overtraining and significantly lower computing resources. Due to single matrix operation, it leads to extremely efficient computation.

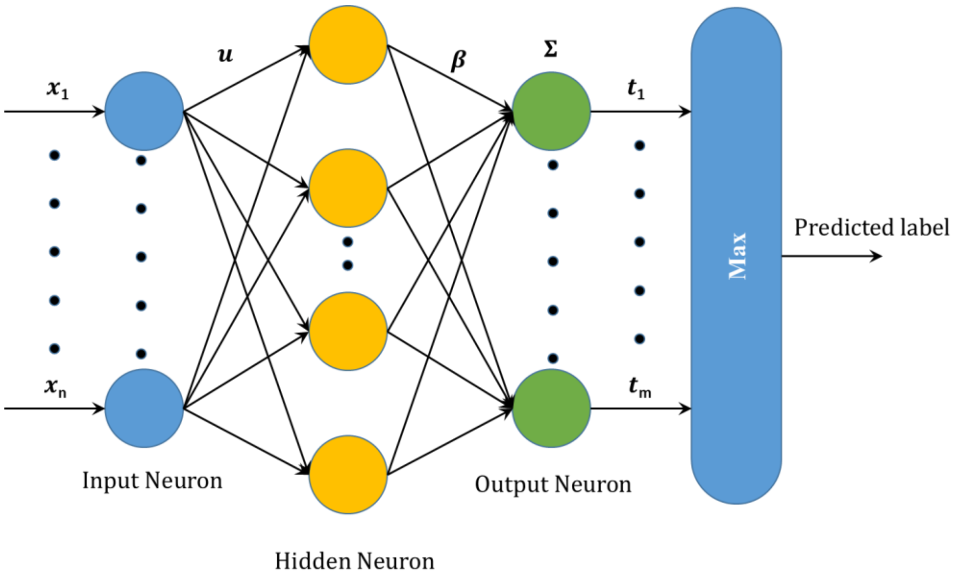


Figure 5.9: Schematics of ELM operation

5.3 Experimentation and Results

In the validation of classification performance on the novel architecture, heartwave from 52 individuals are used. Of the 52 individuals, 5 individuals have heartwaves which are non-healthy. There are 27 individual whose signals have been subjected to extreme physical duress causing maximum morphological variation.

5.3.1 Classification with Anomaly Heartwave

Heartwave based biometric methodologies have been actively tested on individual with normal heartwave signals. Conversely, there are also work in classifying unhealthy heartwave signals into different categories [127, 128]. However, there has been no investigation on the possibility of using individual anomaly heartwave as an authentication mode. For some individuals, the anomaly heartwave signature occurs in every heartwave such as the extended T-Peak. See Figure 5.10 for illustration. For others, the anomaly can be due to the abnormal R-R interval where the period between R peaks are not consistent. See Figure 5.11 for illustration. Of the 52 individuals, there are 5 individuals with anomaly heartwave signals.

The proposed architecture robustness is tested with a mixture of healthy and unhealthy heartwave signals. Importantly, this robustness test determines the possibility of individual with anomaly signal having uniqueness characteristic to support authentication. Similar to the normal healthy signal, anomaly heartwave is tested using 30% of training data with unique 70% data as testing.

In the testing, the proposed architecture is tested on two separate dataset in which one of the dataset contains normal heartwave and the second dataset contains anomaly heartwave signal. Lastly, a third dataset containing all heartwave signal is generated. The performance results are tabulated in Figure 5.12.

From the test performance as shown in Figure 5.12, misclassification rate of 0.8% is achieved on dataset contained healthy heartwave signal. A misclassification rate of 1.8% is achieved on dataset that contains anomaly heartwave signal. The dropped in performance is likely due to the limited data available for training. In combining all heartwave signal into a single dataset, the overall performance is a misclassification rate of 1.1%. In contrast to a statistical method of using Gaussian Mixture Model with Hidden Markov Model (GMM-HMM), the misclassification rate achieved is at 25% under similar proportional of training and testing data. Prior to the proposed architecture, the author proposed a statistical based architecture that uses the characteristic strength of GMM and HMM for classification. For each individual, the dataset belong to a single individual is clustered using GMM. This clustering of individual dataset using GMM allows a user representative model to be developed. The GMM is subsequently used by HMM to develop into a user specific HMM model to support classification.

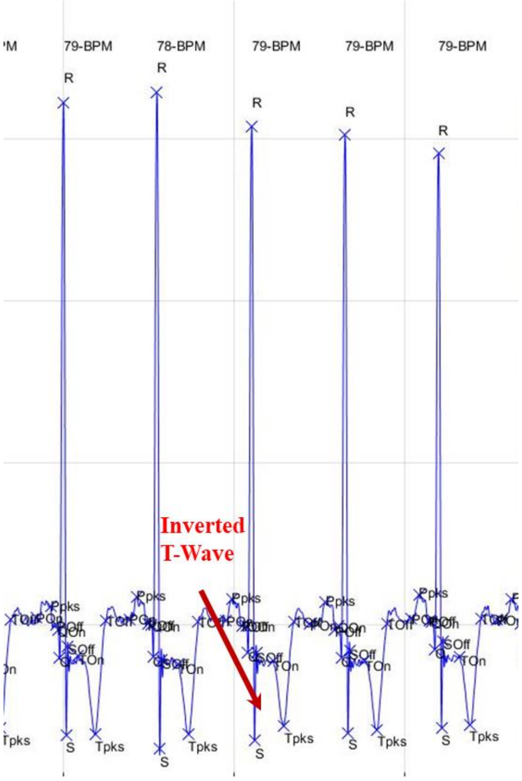


Figure 5.10: Anomaly signal with inverted T-Wave that occurs in every heartwave

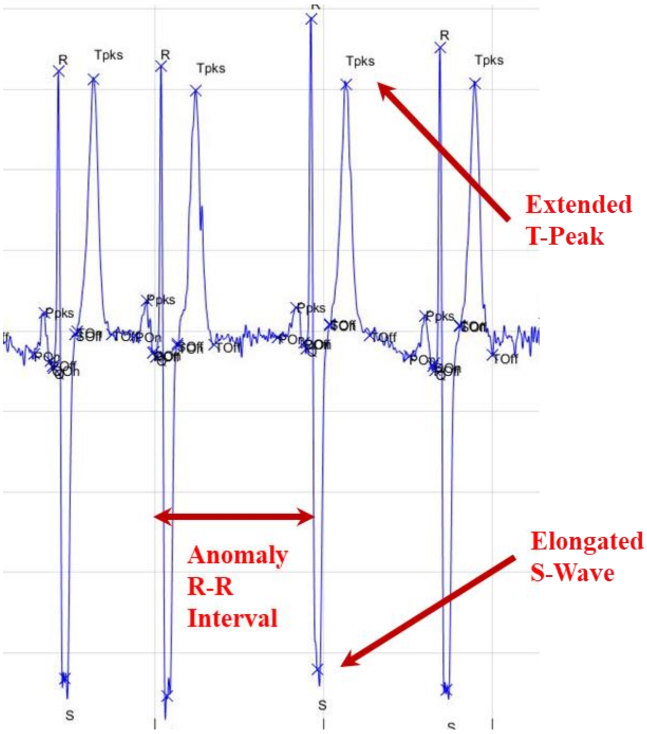


Figure 5.11: Anomaly signal with extended T-Peak and inconsistent R-R interval

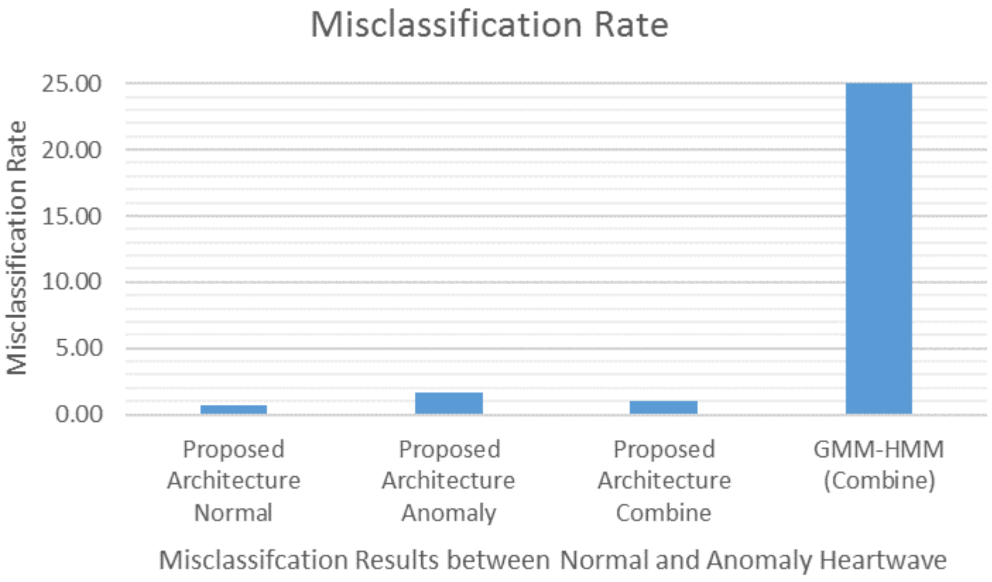


Figure 5.12: Proposed architecture performance on normal and anomaly heartwave signal

5.3.2 Comparison of Single DBN vs Ensemble-DBN

The proposed architecture uses 3 signal DBNs and combined the outputs from each DBN before transferring to the upper module of the proposed architecture consisting of single DBN.

There are reported works that use Deep Learning in biometric authentication and only single DBN is adopted. Primarily, the dataset used in their works are homogenous and most importantly, heartwave signal is acquired under non-physical duress.

In the process of DBN configuration, the hyper parameters consisting of hidden nodes and layers are used to determine the optimized DBN configurations. Considering the dataset that consists of heartwave signals acquired under physical duress, the Signal-to-Noise-Ratio (SNR) of heartwave signals between individual varies. This is primarily due to the placement of the electrodes and motion artifact. Figure 5.13 illustrates a low SNR of a noisy heartwave signal. Further investigations reveal that for heartwave signal with high SNR, the DBN configuration with lower number of hidden nodes performs well in classification, in excess of 98% accuracy. Conversely, for heartwave signal with low SNR, DBN with higher number of hidden nodes performs better in classification. To further complicates the process the classification process, the SNR of most individual heartwave signal varies at different heartrates. During the heartwave signal acquisition process, for most individual, at the early stage of the treadmill testing where the heartrate is near resting heartrate, the signal has high SNR. With increasing intensity of the treadmill, the SNR deteriorates at increasing heartrates. This concludes that having multiple DBNs is necessary to manage the varying SNR of individual heartwave signal.

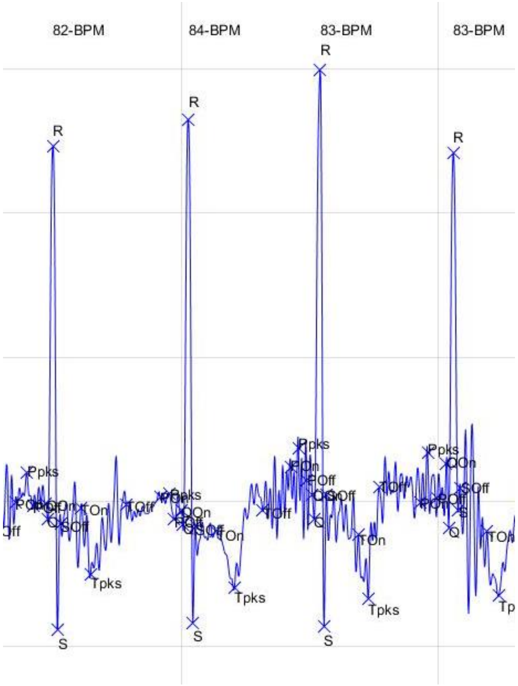


Figure 5.13: Extreme noisy signal with low SNR

In this investigation, the performance of the proposed configuration for each of the three DBNs is compared. The results are shown in Figure 5.14. For single DBN with configuration of 500x500x500, the classification accuracy is in excess of 91% as compared to DBNs with higher number of hidden nodes. This is primarily because of the large number of heartwaves having higher SNR at lower heartrate. Leveraging on the strength of each of the single DBN, the three outputs are combined through the adoption of MSE. The MSE works to determine views with higher significant strength which output a weight for each of the views. The greater the significance of the view, the higher the weight. This results in the ability of the proposed architecture to achieve a classification accuracy of 98.3%.

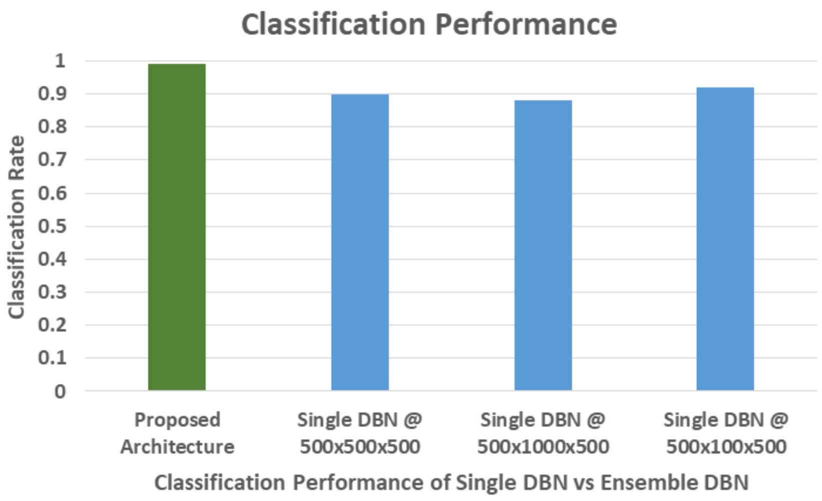


Figure 5.14: Classification performance of various standalone DBN against proposed architecture

5.3.3 Classification Performance on Proportion of Training Data with Proposed Architecture

In some of the reported works, the proportion of training data against testing data is benchmarked from 70% to 90% [92, 128]. While having the 70% of the data may seems appropriate, it is necessary to explore the strength of Deep Learning with the aim to reduce the training data required. Importantly, it is of the opined that the apportioning of 70% of data for training is still excessive.

The investigation starts with the varying percentage of training data. In addition, the investigation also examines the performance of the three single DBNs against the proposed architecture. The results of the classification are summarized in Figure 5.15. From Figure 5.15, it is evident that the conventional apportioning of 70% of data for training is reasonable. At 70% of data for training, the classification performance easily achieved an accuracy rate in excess of 95%. With decreasing training data, the performance deteriorates rapidly.

It is noted that for the proposed architecture, the classification performance remains consistent at about 98.3% even at 30% training data. This is in stark contrast to the performance of individual DBNs. As discussed in Section 5.3.2, where due to variable heartrate, leads to variable signal-to-noise ratio. Each of the DBN has limited capability to capture all the morphological variations of heartwave signal. For heartwave signal with high SNR, the DBN configuration with lower number of hidden nodes performs differently in contrast to heartwave signal with low SNR. Importantly, this reinforces the strength in importance of incorporating MSE into the architecture. As explained in earlier section, MSE determines the significant view contribution and output the appropriate weight for respective view. This adoption of MSE has contributed significantly to the classification performance under low training data.

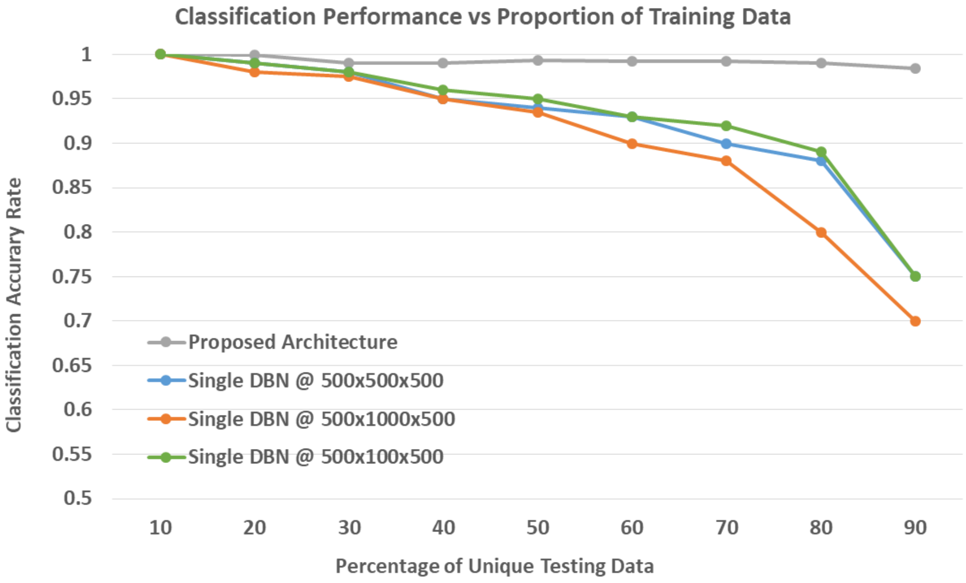


Figure 5.15: Classification performance under different proportion of training data

5.3.4 Comparison of proposed architecture with statistical methodology

The performance of the proposed architecture is evaluated and compared against other statistical methodologies. The work in [129] uses a combination of fiducial, non-fiducial data of heartwave signal and inertial sensor parameter and achieved a True

Positive Rate of approximately 80% with False Negative Rate of 12%. A commonly adopted methodology is the generative modelling such as Gaussian Mixture Modelling and Hidden Markov Modelling which are used to support data clustering and classification. Hence, the proposed architecture is compared against generative modelling of Gaussian Mixture Model with Hidden Markv Model. The GMM-HMM method achieved a True Positive Rate of approximately 90% with False Negative Rate of 10%. In comparison of proposed architecture, the True Positive Rate achieved is 98% with False Negative Rate of under 2%. See Figure 5.16 for comparison. Deep learning methodology offers a highly reliable approach to classification.

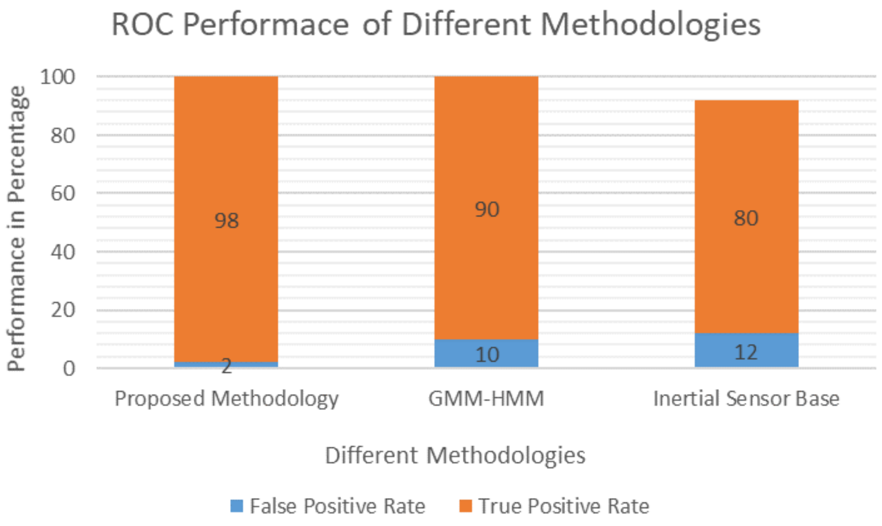


Figure 5.16: Comparison of classification performance between proposed architecture and statistical methods

5.3.5 Performance of Proposed Architecture with Architecture 2

The proposed architecture uses a coefficient factor from MSE to amplify view with significant contribution. This proposed architecture has resulted a classification performance of 98.3%. The expression (5.36) can be expressed into the following form where v_i represents the penultimate layer of DBN i and α_i refers to the multiplier factor for view i from MSE.

$$\Psi_{user} = f_{MSE}(v_i\alpha_i + \cdots + v_m\alpha_m) \tag{5.41}$$

An alternative architecture (termed as Architecture 2) has been investigated. In Architecture 2, in contrast to the proposed architecture, all the outputs of the penultimate layers are arithmetically sum and average. The output of the MSE, Y, which contains the embedded low-dimensional complementary property from all the views is concatenate with the averaged penultimate layers. The input to the stacked DBN can be expressed as shown in (5.42). The eventual dataset, Ψ_{user} from Architecture 2 is of higher dimension than proposed architecture due to the concatenation. See Figure 5.17 for schematic of proposed Architecture 2.

$$\Psi_{user} = f_{MSE}\left(Concat\left(\frac{v_i + \cdots + v_m}{m}, Y\right)\right) \tag{5.42}$$

Figure 5.18 shows the performance similarity between the proposed architecture and the alternative architecture. The performance between the proposed and alternative are similar at approximately around 98.2%. Interesting, the architecture 2 is able to perform similarly with reduce training sample size.

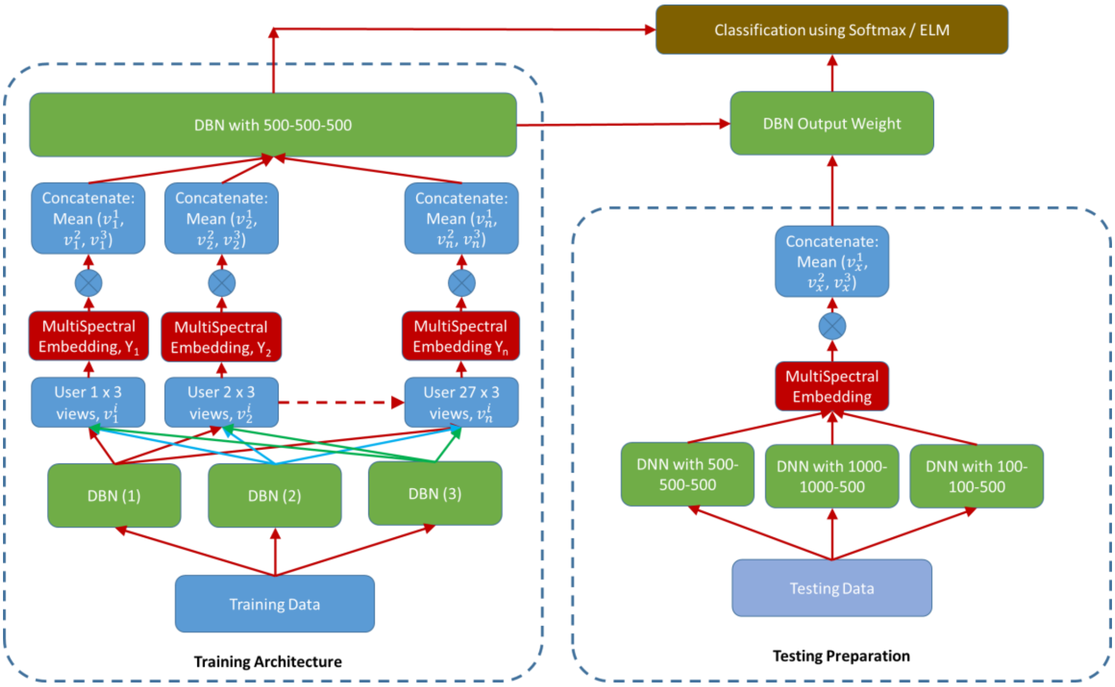


Figure 5.17: Illustration of the proposed architecture 2

The performance of architecture 2 reaffirms the strength of MSE. The difference between expression (5.41) and (5.42) is the used of MSE output. Proposed architecture uses multiplier factor α, to amplify views with significant contribution. Architecture 2 uses Y, which is the embedded low dimensional data structure complimentary of the three single DBNs. Although the multiplier factor is not utilized, the embedded low dimension structure Y has already been encoded with view containing significant property. Thus the embedded low dimensional structure plays a key role in the higher module classification.

The results from the Architecture 2 is tabulated into a Confusion Table as shown in Table 5.1 and evaluated against GMM-HMM methodology in particular to Table 4.3. To ensure a consistent comparison, the Architecture 2 is trained with 30% Training Data and tested on 70% unique Testing Data.

In Table 4.3, User 312 has an Error of Commission at 16.68% and User 313 with an Error of Omission at 7.18%. Using Architecture 2, User 312 Error of Commission improved by more than a factor at 0.73% and User 313 has an Error of Omission at 1.32%.

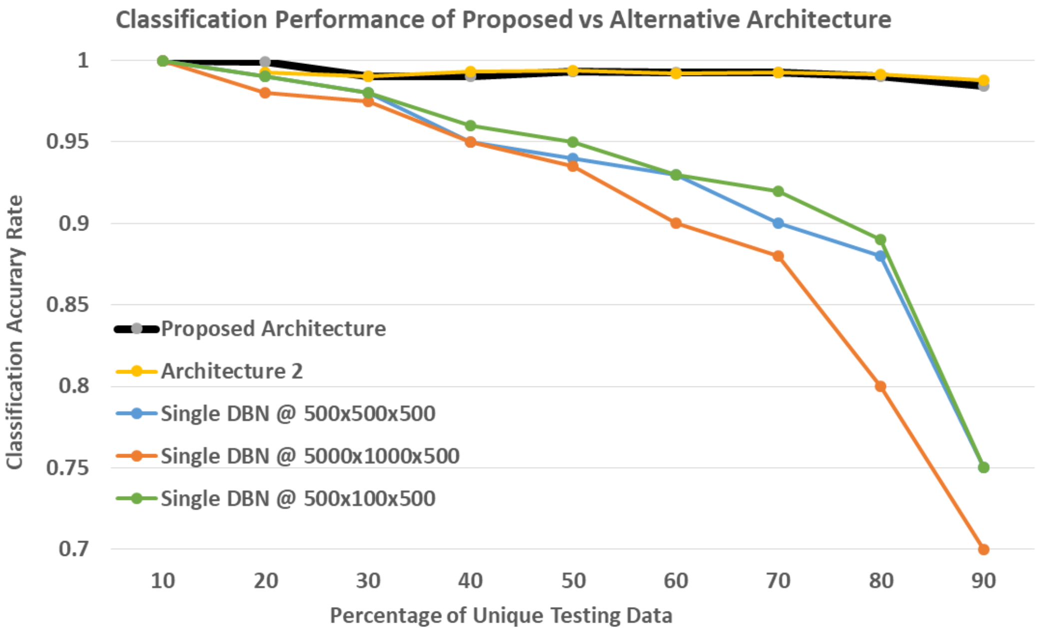


Figure 5.18: Comparison of classification performance between proposed architecture and architecture 2 (alternative)

The vast improvement exhibited the strength of neural network based methodology. The implementation of Architecture 2 is almost similar to this thesis proposed architecture except that the outputs of the penultimate layers are arithmetically summed and averaged. The output of the MSE, Y, which contains the embedded low-dimensional complementary property from all the views is concatenate with the averaged penultimate layers before feeding to stacked DBN as expressed in (5.42).

Importantly, in GMM-HMM methodology, only temporal data are used to determine the discriminating feature of individual. In Architecture 2, the heartwave is individual extracted and the full heartwave profile containing amplitude and temporal details are used for training and discriminating. Equally important, the use of MultiView Spectra Embedding (MSE) to combine the multiple outputs from ensemble-DBN into a single structure has enabled significant discriminating features from multiple views to play a significant role in discriminating.

Table 5.1: Confusion Table for Architecture 2 with 52 users.

		User 300	User 301	User 302	User 303	User 304	User 306	User 307	User 308	User 309	User 310	User 311	User 312	User 313	User 314	User 316	User 317	User 319	User 320	User 321
Actual	User 300	1563	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
	User 301	0	1450	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
	User 302	0	0	1253	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 303	0	0	0	1777	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 304	0	0	0	0	1033	0	1	0	0	0	0	0	0	0	0	0	2	0	0
	User 306	0	0	0	0	0	3989	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 307	0	0	0	0	0	0	1405	0	0	0	0	0	0	0	0	0	0	1	0
	User 308	0	0	0	0	0	0	0	1281	0	0	0	0	0	0	0	0	0	0	0
	User 309	0	0	0	0	0	0	0	0	3234	0	0	0	0	0	1	0	1	0	0
	User 310	0	0	0	0	0	0	0	0	0	1509	0	0	0	0	1	1	0	0	0
	User 311	0	0	0	0	0	0	0	0	1	0	1807	0	4	2	0	2	2	0	0
	User 312	0	1	1	0	0	0	0	0	0	0	0	1362	0	0	0	0	12	0	0
	User 313	0	0	0	0	0	0	0	0	1	1	5	1	1650	0	0	4	3	0	0
	User 314	0	0	0	0	0	0	0	0	0	0	0	0	0	1241	0	0	0	0	0
	User 316	0	0	1	0	0	0	0	0	0	0	0	0	2	0	2039	0	0	0	0
	User 317	0	0	1	0	0	2	0	0	1	0	3	0	0	0	0	1663	0	12	0
	User 319	2	0	0	1	0	1	0	0	0	0	0	4	0	4	1	0	1444	1	0
	User 320	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	13	2	1873	0
	User 321	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1275
	User 322	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	2	0	0
	User 323	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
	User 324	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 325	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 326	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	User 327	0	0	0	1	0	0	0	0	0	0	0	0	2	0	1	0	3	0	0
	User 328	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	User 329	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 330	1	0	2	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0
	User 331	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 332	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	User 333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	User 334	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 335	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 336	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0
	User 337	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0
	User 338	0	0	3	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0
	User 339	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	User 340	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 341	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 342	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	4	1	2	0
	User 343	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0
	User 344	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 345	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
	User 346	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
	User 347	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 348	0	0	0	0	0	0	0	0	1	0	0	1	0	1	4	0	2	0	0
	User 349	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 350	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
	User 351	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	User 352	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	User 353	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	User 354	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Total	1567	1454	1264	1779	1034	3994	1409	1287	3238	1514	1835	**1372**	1659	1252	2049	1688	1481	1894	1276
	Err Comm	4	4	11	2	1	5	4	6	4	5	28	**10**	9	11	10	25	37	21	1
	% Err Comm	0.26	0.28	0.87	0.11	0.10	0.13	0.28	0.47	0.12	0.33	1.53	**0.73**	0.54	0.88	0.49	1.48	2.50	1.11	0.08

	User 354	Total	Err Omis	% Err Omis
User 300	0	1568	**5**	**0.32**
User 301	0	1455	**5**	**0.34**
User 302	0	1253	**0**	**0.00**
User 303	0	1777	**0**	**0.00**
User 304	0	1036	**3**	**0.29**
User 306	0	3990	**1**	**0.03**
User 307	0	1407	**2**	**0.14**
User 308	0	1281	**0**	**0.00**
User 309	0	3241	**7**	**0.22**
User 310	0	1511	**2**	**0.13**
User 311	0	1826	**19**	**1.04**
User 312	0	1379	**17**	**1.23**
User 313	0	1672	**22**	**1.32**
User 314	0	1244	**3**	**0.24**
User 316	0	2044	**5**	**0.24**
User 317	0	1686	**23**	**1.36**
User 319	0	1477	**33**	**2.23**
User 320	0	1910	**37**	**1.94**
User 321	0	1281	**6**	**0.47**
User 322	0	931	**8**	**0.86**
User 323	0	838	**3**	**0.36**
User 324	0	706	**6**	**0.85**
User 325	0	1414	**3**	**0.21**
User 326	0	952	**4**	**0.42**
User 327	0	1266	**9**	**0.71**
User 328	0	1218	**4**	**0.33**
User 329	0	1343	**2**	**0.15**
User 330	0	1295	**12**	**0.93**
User 331	0	1021	**0**	**0.00**
User 332	0	1092	**3**	**0.27**
User 333	0	916	**1**	**0.11**
User 334	3	1134	**6**	**0.53**
User 335	0	1211	**0**	**0.00**
User 336	0	1603	**4**	**0.25**
User 337	0	1335	**8**	**0.60**
User 338	0	1545	**28**	**1.81**
User 339	0	1225	**6**	**0.49**
User 340	0	965	**0**	**0.00**
User 341	0	1400	**0**	**0.00**
User 342	0	1638	**12**	**0.73**
User 343	0	1007	**3**	**0.30**
User 344	0	1148	**7**	**0.61**
User 345	0	1182	**2**	**0.17**
User 346	0	1015	**3**	**0.30**
User 347	0	1798	**0**	**0.00**
User 348	0	1168	**13**	**1.11**
User 349	0	1169	**2**	**0.17**
User 350	0	1155	**7**	**0.61**
User 351	0	1400	**2**	**0.14**
User 352	0	1169	**1**	**0.09**
User 353	0	1043	**4**	**0.38**
User 354	1159	1160	**1**	**0.09**
Total	1162			
Err Comm	3			
% Err Comm	0.26			

5.4 Summary

The proposed architecture in the heartwave based biometric authentication exceeded statistical methodology. Importantly, classification on individual heartwave with intense varying heartrates which causes signification morphologically variations has been tested by the proposed architecture. The proposed architecture has shown the ability to identify individuals comprising of normal and abnormal heartwave signals with high level of reliability. Architecturally, the ensemble-DBN is necessary to enable feature extractions under different morphological variations. The incorporation of MSE has enabled views with significant features with greater biasness in aggregated inputs to stacked DBN module.

CHAPTER 6

CONCLUSION OF THESIS

The work in this thesis have concluded the objectives of the research on the development and investigation of new methodologies to enable a better and reliable performance of a heartwave based biometric classification. More importantly, the proposed solutions and architectures are robust and resilient to heartwave morphological variation and heartrate variation.

Chapter 2 provided an overview of the current states-of-the-art in the domains of heartwave segmentation and extraction, and identifying discriminating features for classification. The various algorithms and methodologies are suited for different type of signals in different situations based on their limitations and constraints. In particular to the heartwave segmentation and extraction of features, none of the reported work have been attempt on heartwave signal under elevated heartrate. With regards to the classification, the review of the reported work provided an in-depth understanding of the limitations in performing heartwave based biometric in particular to heartwave signal under extreme morphological variation. This fact is the motivation of this thesis, which commits to develop new algorithms and methodologies to address extraction and classification issues caused by elevated heartrate.

In Chapter 3, a methodology consisting of Discrete Waveform Transformation integrated with heartrate dependent parameters: PR-Interval and QT-Interval has been proposed and tested to perform 11 heartwave features identification and extraction. In total, more than 63,000 heartwaves have been extracted from 27 individuals. The dataset of the 27 individuals contains heartwave signal acquired during extreme exercise duress. Specifically, it achieved 100% detection accuracy for R-Peak and an average of 98% for the other 10 heartwave features. The use of heartrate dependent windows for PR Interval and QT Interval to support detection of wave components before and after R-Peak has achieved superior performance as compared to fixed window length. In addition, the same algorithm has been able to extract heartwave features from anomaly heartwave signal. From the database of 27 users, there are 5 individuals with anomaly heartwave signals which include signals with extended T-Peak, inverted T-Wave and irregular R-R interval. The ability to extract characteristic features with high accuracy and high sensitivity concluded the ability of the proposed solution to extract heartwave features reliably and accurately under highly variated heartrate.

In chapter 4, a novel architecture consisting of statistical based methodology of Gaussian Mixture Model integrated with Hidden Markov Model (GMM-HMM) aided with discriminating criteria of user specific thresholding score and heartrate range has been tested. The presentation of heartwave morphological variations due to heartrate has prompted a generative modelling methodology to model the joint distribution of data on individual dataset. The GMM generated model is unique to individual and the integration of Hidden Markov Model is implemented to perform discrimination of individual. Through this development, it is observed the individual matching score via loglikelihood is a linear behavior against individual heartrate. In addition, every individual has its own unique heartrate range. These understanding lead to the affirmation that individual discriminating features are not static and it vary according to its respective heartrate at the instance. To cater for the variation, the two imposing criteria, individual heartrate range and individual unique thresholding behavior are implemented. The proposed architecture achieved an accuracy of 89% from 27 individuals. More importantly, all the heartwave signals contain heartwave with extreme heartwave morphological variations. Comparing the novel proposed architecture against commonly adopted methodology of Linear Discriminant Analysis (LDA) with the same dataset, LDA achieved an accuracy of 78% with a False Positive Rate of 25%. The proposed architecture doubled the performance and achieved a False Positive Rate of 11%. This development and investigation conclude the need for proposed methodology to account for varying discriminating features. Unless an individual has no varying heartrate, methodology with fixed and static discriminating feature will more likely result in poor classification accuracy. The proposed methodology has achieved its intending objective to develop a classification model and methodology to perform biometric classification which is resilient to individual dynamic and varying heartrate.

In Chapter 5, a novel architecture categorized under neural network, has been developed and investigated. The architecture consists of an ensemble of Deep Belief Networks (DBN) connected to a module of MultiView Spectral Embedding (MSE) to combine the multiple output from ensemble-DBN into a single structure that contains significant discriminating features from multiple views. The single view structure is further input to a stacked DBN to perform classification and output via an effective and resource efficient method of Extreme Learning Machine. The performance of the proposed architecture is impressive. The proposed neural network of Deep Ensemble Architecture, is tested on 52 individuals consisting of 22 individuals with extreme heartwave morphological variations, 25 individuals whose heartwave are acquired under resting condition and 5 individuals under extreme heartwave morphological variation with anomaly heartwave signals. The proposed architecture has an accuracy of 98.9% . Equally important, the proposed architecture is capable of achieving 98.3% accuracy with a limited portion of the training data at 30%. This result is a vast improvement against most reported work where the 60% to 80% of the available data has been apportioned for training. This investigation leads to the understanding that the increase in nodes in DBN is not necessary an advantage for highly variated signal for low signal-to-noise-ratio (SNR). DBN with lower number of nodes will results in under-fit for highly variated signals with low SNR. Conversely, DBN with high dimensional units suffers from over-fit for highly variated signals with low SNR. MSE plays a key role in identifying complementary property on views with higher order of significant. The development of the proposed neural network architecture has achieved its intended objective to develop a classification model and methodology to perform biometric classification which is resilient to individual dynamic and varying heartrate at lower proportion of training data.

6.1 Future Works

Two novel architectures, statistical and neural network methodologies, have been proposed to address heartwave based biometric constraints dues to varying heartrate and the results demonstrated promising performance. Heartwave based biometric has vast potentials to support the current and future digital age services such as IoT, cloud services and big data service. Before the solution is ready for adoption, there are works which require further exploration.

6.1.1 Improvement to Deep Neural Network based Architecture

Although the performance for deep ensemble-DBN architecture has been stellar, there exist an opportunity for further improvement of the architecture. In the current setup, there is a need for data preparation prior to the initialization of ensemble-DBN. The data structure requires the heartwave to be individual extracted and compiled into stacked of single column of single heartwave via Chapter 3 proposed Heartrate dependent Discrete Wavelet Transformation solution. With DBN discrimination feature extraction capability and to eliminate the single heartwave segmentation process, it is recommended to explore the possibility of incorporating another layer of DBN primarily to identify the segment of a single whole heartwave from a sequence of input source that contains a series of heartwave signals.

6.1.2 Lack of Heartwave Data under Extreme Physical Duress

The current work and testing of the methodologies is limited by lack of data whose heartwave signals have been subjected to treadmill testing. The current data are extracted from Physionet –ST Change and it contains only 27 users. More datasets are required to stretch and stress the proposed architectures of heartwave segmentation under variable heartrates and heartwave based classification.

6.1.3 Multi-Modality Biometric Solution

With the emergence of wearable devices that is capable to extend the measurement to SPO2, perspiration, blood pressure, heartrate and motion sensor, it is possible to incorporate these measurements as part of the discriminating features for either classification or monitoring of individual wellness. The proposed deep neural network architecture has the potential to expand the ensemble-DBN to cater to additional dimensions of measurement parameters. The module of MSE has proven to be useful to identify significant feature of interest and combining the complementary properties from Multiview into a single structure view.

REFERENCES

[1] N. A. Albahbooh and P. Bours, "A Mobile Phone Device as a Biometrics Authentication Method for an ATM Terminal," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015, pp. 2017-2024.

[2] K. T. Chui, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, "An Accurate ECG-Based Transportation Safety Drowsiness Detection Scheme," *IEEE Transactions on Industrial Informatics,* vol. 12, no. 4, pp. 1438-1452, 2016.

[3] G. Yang *et al.*, "A Health-IoT Platform Based on the Integration of Intelligent Packaging, Unobtrusive Bio-Sensor, and Intelligent Medicine Box," *IEEE Transactions on Industrial Informatics,* vol. 10, no. 4, pp. 2180-2191, 2014.

[4] L. Hamid, "Biometric technology: not a password replacement, but a complement," *Biometric Technology Today,* vol. 2015, no. 6, pp. 7-10, 2015/06/01/ 2015.

[5] D. R. Ibrahim, A. A. Tamimi, and A. M. Abdalla, "Performance analysis of biometric recognition modalities," in *2017 8th International Conference on Information Technology (ICIT)*, 2017, pp. 980-984.

[6] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters,* vol. 79, pp. 80-105, 2016/08/01/ 2016.

[7] A. S. Raju and V. Udayashankara, "Biometric person authentication: A review," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 2014, pp. 575-580.

[8] N. Sivasankari and A. Muthukumar, "A review on recent techniques in multimodal biometrics," in *2016 International Conference on Computer Communication and Informatics (ICCCI)*, 2016, pp. 1-11.

[9] Y. Sun, M. Zhang, Z. Sun, and T. Tan, "Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 2, pp. 332-351, 2018.

[10] H. Thakuria *et al.*, "A comparative study of vein pattern recognition for biometrie authentication," in *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2017, pp. 689-694.

[11] Y. Si, J. Mei, and H. Gao, "Novel Approaches to Improve Robustness, Accuracy and Rapidity of Iris Recognition Systems," *IEEE Transactions on Industrial Informatics,* vol. 8, no. 1, pp. 110-117, 2012.

[12] S. Gaur, V.A.Shah, and M. Thakker, "Biometric Recognition Techniques: A Review," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* vol. 1, no. 4, pp. 282-290, October 2012.

[13] J. Yang, Ed. *Biometrics*. InTech,, 2011, p. 266.

[14] T. Burghardt, *A brief review of biometric identification*. Technical Report, Visual Information Laboratory, Bristol, 2012.

[15] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognition,* vol. 38, no. 1, pp. 133-142, 2005.

[16] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement,* vol. 50, no. 3, pp. 808-812, 2001.

[17] F. Sufi, I. Khalil, and J. Hu, "ECG-Based Authentication," in *Handbook of Information and Communication Security*, P. Stavroulakis and M. Stamp, Eds.: Springer Berlin Heidelberg, 2010, pp. 309-331.

[18] I. Odinaka, L. Po-Hsiang, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "ECG Biometric Recognition: A Comparative Analysis," *Information Forensics and Security, IEEE Transactions on,* vol. 7, no. 6, pp. 1812-1824, 2012.

[19] G. R. Shaw and P. Savard, "On the detection of QRS variations in the ECG," (in eng), *IEEE Trans Biomed Eng,* vol. 42, no. 7, pp. 736-41, Jul 1995.

[20] L. S. Green, R. L. Lux, C. W. Haws, R. R. Williams, S. C. Hunt, and M. J. Burgess, "Effects of age, sex, and body habitus on QRS and ST-T potential maps of 1100 normal subjects [published erratum appears in Circulation 1986 Oct;74(4):785]," *Circulation,* vol. 71, no. 2, pp. 244-253, 1985.

[21] F. Agrafioti and D. Hatzinakos, "Signal validation for cardiac biometrics," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 1734-1737.

[22] M. Homer, J. M. Irvine, and S. Wendelken, "A model-based approach to human identification using ECG," 2009, vol. 7306, pp. 730625-730625-10.

[23] K. A. Sidek, I. Khalil, and M. Smolen, "ECG biometric recognition in different physiological conditions using robust normalized QRS complexes," in *Computing in Cardiology (CinC), 2012*, 2012, pp. 97-100.

[24] K. Kyeong-Seop, Y. Tae-Ho, L. Jeong-Whan, K. Dong-Jun, and K. Heung-Seo, "A Robust Human Identification by Normalized Time-Domain Features of Electrocardiogram," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2005, pp. 1114-1117.

[25] P. Shing-Tai, W. Yi-Heng, K. Yi-Lan, and C. Hung-Chin, "Heartbeat Recognition from ECG Signals Using Hidden Markov Model with Adaptive Features," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013 14th ACIS International Conference on*, 2013, pp. 586-591.

[26] E. Rabhi and Z. Lachiri, "Biometric personal identification system using the ECG signal," in *Computing in Cardiology Conference (CinC), 2013*, 2013, pp. 507-510.

[27] P. Shing-Tai, H. Tzung-Pei, and C. Hung-Chin, "ECG signal analysis by using Hidden Markov model," in *Fuzzy Theory and it's Applications (iFUZZY), 2012 International Conference on*, 2012, pp. 288-293.

[28] F. Agrafioti and D. Hatzinakos, "ECG biometric analysis in cardiac irregularity conditions," *Signal, Image and Video Processing,* vol. 3, no. 4, pp. 329-343, 2008.

[29] J. Carlson, R. Johansson, and S. B. Olsson, "Classification of electrocardiographic P-wave morphology," *Biomedical Engineering, IEEE Transactions on,* vol. 48, no. 4, pp. 401-405, 2001.

[30] J. Hill and A. Timmis, "ABC of clinical electrocardiography. Exercise tolerance testing," *BMJ,* vol. 324, no. 7345, pp. 1084-1087, 05/04/ 2002.

[31] G. F. Fletcher, G. Balady, V. F. Froelicher, L. H. Hartley, W. L. Haskell, and M. L. Pollock,

"Exercise standards. A statement for healthcare professionals from the American Heart Association. Writing Group," (in eng), *Circulation,* vol. 91, no. 2, pp. 580-615, Jan 15 1995.

[32] M. Pilhall, M. Riha, and S. Jern, "Exercise-induced QRS changes in healthy men and women: a multivariate analysis on their relation to background data and exercise performance," (in eng), *Eur Heart J,* vol. 13, no. 10, pp. 1316-24, Oct 1992.

[33] M. L. Simoons and P. G. Hugenholtz, "Gradual changes of ECG waveform during and after exercise in normal subjects," *Circulation,* vol. 52, no. 4, pp. 570-577, 1975.

[34] M. Malik, K. Hnatkova, M. Sisakova, and G. Schmidt, "Subject-specific heart rate dependency of electrocardiographic QT, PQ, and QRS intervals," *Journal of Electrocardiology,* vol. 41, no. 6, pp. 491-497, 11// 2008.

[35] P. N. G. Yu, R. A. Bruce, F. W. Lovejoy, and M. E. McDowell, "Variations in Electrocardiographic Responses during Exercise: Studies of Normal Subjects under Unusual Stresses and of Patients with Cardiopulmonary Diseases," *Circulation,* vol. 3, no. 3, pp. 368-376, 1951.

[36] V. N. Batchvarov, G. Bortolan, and I. I. Christov, "Effect of heart rate and body position on the complexity of the QRS and T wave in healthy subjects," in *Computers in Cardiology, 2008*, 2008, pp. 225-228.

[37] J. O. Ogedengbe, A. B. Adelaiye, and O. V. Kolawole, "Effects of exercise on PR intervals, QRS durations and QTC intervals in male and female students of University of Abuja," (in eng), *J Pak Med Assoc,* vol. 62, no. 3, pp. 273-5, Mar 2012.

[38] P. M. Rautaharju, S. Punsar, H. Blackburn, J. Warren, and A. Menotti, "Waveform Patterns in Frank-Lead Rest and Exercise Electrocardiograms of Healthy Elderly Men," *Circulation,* vol. 48, no. 3, pp. 541-548, 1973.

[39] J. U. Lee, K. S. Kim, J. H. Kim, H. K. Lim, B. H. Lee, and C. K. Lee, "PR interval behavior during exercise stress test," (in eng), *Korean J Intern Med,* vol. 10, no. 2, pp. 137-42, Jul 1995.

[40] S. M. Al-Khatib, N. M. LaPointe, J. M. Kramer, and R. M. Califf, "What clinicians should know about the QT interval," (in eng), *JAMA,* vol. 289, no. 16, pp. 2120-7, Apr 23-30 2003.

[41] C. E. Garnett *et al.*, "Methodologies to characterize the QT/corrected QT interval in the presence of drug-induced heart rate changes or other autonomic effects," *Am Heart J,* vol. 163, no. 6, pp. 912-30, Jun 2012.

[42] G. Kaleschke *et al.*, "Prospective, multicentre validation of a simple, patient-operated electrocardiographic system for the detection of arrhythmias and electrocardiographic changes," *Europace,* vol. 11, no. 10, pp. 1362-8, Oct 2009.

[43] M. M. Tawfik, H. Selim, and T. Kamal, "Human identification using time normalized QT signal and the QRS complex of the ECG," in *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, 2010, pp. 755-759.

[44] H. Qiu, G. L. Bird, L. Qu, V. L. Vetter, and P. S. White, "Evaluation of QT interval correction methods in normal pediatric resting ECGs," in *Computers in Cardiology, 2007*, 2007, pp. 431-434.

[45] B. Hosmane, C. Locke, and D. Morris, "QT Interval Correction for Heart Rate.pdf>," *The Journal of Applied Research,* vol. 6, no. 4, pp. 288-299, 2006.

[46] M. Li and X. Li, "Verification based ECG biometrics with cardiac irregular conditions using heartbeat level and segment level information fusion," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence,* pp. 3769-3773, 2014.

[47] I. Odinaka *et al.*, "ECG biometrics: A robust short-time frequency analysis," *Proc. 2010 IEEE Int. Workshop Information Forensics and Security (WIFS), Dec. 2010,* pp. 1-6, 2010.

[48] M. Li and S. Narayanan, "Robust ECG biometrics by fusing temporal and cepstral information," *Proc. 2010 20th Int. Conf. Pattern Recognition (ICPR), Aug, 2010,* pp. 1326-1329, 2010.

[49] J. M. Irvine and S. A. Israel, "A Sequential Procedure for Individual Identity Verification Using ECG," *EURASIP Journal on Advances in Signal Processing,* journal article vol. 2009, no. 1, p. 243215, May 11 2009.

[50] Z. Zhang and D.Wei, "A new ECG identification method using Bayes' theorem," *Proc. IEEE Region 10 Conf.,,* pp. 1-4, 2006.

[51] K. Plataniotis, D. Hatzinakos, and J. K. M. Lee, *ECG Biometric Recognition Without Fiducial Detection*. 2006, pp. 1-6.

[52] M. N. Dar, M. U. Akram, A. Shaukat, and M. A. Khan, "ECG Based Biometric Identification for Population with Normal and Cardiac Anomalies Using Hybrid HRV and DWT Features," *2015 5th International Conference on IT Convergence and Security (ICITCS), Kuala Lumpur,* pp. 1-5, 2015.

[53] Z. Wang and Y. Zhang, "Research on ECG Biometric in Cardiac Irregularity Conditions," *2014 International Conference on Medical Biometrics, Shenzhen,* pp. 157-163, 2014.

[54] S. L. Lin, C. K. Chen, C. L. Lin, W. C. Yang, and C. T. Chiang, "Individual identification based on chaotic electrocardiogram signals during muscular exercise," *IET Biometrics,* vol. 3, no. 4, pp. 257-266, 2014.

[55] C. Bhagavatula *et al.*, "Automatic segmentation of cardiosynchronous waveforms using cepstral analysis and continuous wavelet transforms," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 2045-2048.

[56] D. Sadhukhan and M. Mitra, "Detection of ECG characteristic features using slope thresholding and relative magnitude comparison," in *Emerging Applications of Information Technology (EAIT), 2012 Third International Conference on*, 2012, pp. 122-126.

[57] S. Pal and M. Mitra, "Detection of ECG characteristic points using Multiresolution Wavelet Analysis based Selective Coefficient Method," *Measurement,* vol. 43, no. 2, pp. 255-261, 2010.

[58] L. Cuiwei, Z. Chongxun, and T. Changfeng, "Detection of ECG characteristic points using wavelet transforms," *IEEE Transactions on Biomedical Engineering,* vol. 42, no. 1, pp. 21-28, 1995.

[59] S. Z. Mahmoodabadi, A. Ahmadian, and M. D. Abolhasani, "ECG feature extraction using daubechies wavelets," *5th IASTED International Conference Visualization, Imaging, and Image Processing,* pp. 343-348, September 7-9 2005.

[60] P. Zicari, A. Amira, G. Fischer, and J. McLaughlin, "An embedded system for on field testing of human identification using ECG biometric," in *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, 2012, pp. 65-70.

[61] S. Z. Mahmoodabadi, A. Ahmadian, M. Abolhasani, P. Babyn, and J. Alirezaie, "A fast expert system for electrocardiogram arrhythmia detection," *Expert Systems,* vol. 27, no. 3, pp. 180-200, 2010.

[62] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 11, no. 7, pp. 674-693, 1989.

[63] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet Distance Measure for Person

Identification Using Electrocardiograms," *Instrumentation and Measurement, IEEE Transactions on,* vol. 57, no. 2, pp. 248-253, 2008.

[64] C. L. P. Lim, W. L. Woo, and S. S. Dlay, *Enhanced Wavelet Transformation for Feature Extraction in Highly Variated ECG Signal* (IET Conference Proceedings). Institution of Engineering and Technology, 2015, pp. 6 .-6 .

[65] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of Human Electrocardiogram for Biometric Recognition," *EURASIP Journal on Advances in Signal Processing,* journal article vol. 2008, no. 1, p. 148658, September 19 2007.

[66] D. M, A. Jain, M. S. C, A. Melkot, and M. A. V, "ECG based authentication using Autocorrelation and Artificial Neural Networks," *2016 International Conference on Computing, Analytics and Security Trends (CAST), Pune,* pp. 238-243, 2016.

[67] Y.Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP J. Adv. Signal Process,* vol. 2008, no. 19, pp. 19:1–19:11, 2008.

[68] F. Agrafioti and D. Hatzinakos, "Fusion of ECG sources for human identification," *Proc. 3rd Int. Symp. Communications, Control and Signal Processing (ISCCSP 2008), St. Julians, Malta,* 2008.

[69] A. Jovic and N. Bogunovic, *Feature Extraction for ECG Time-Series Mining Based on Chaos Theory*. 2007, pp. 63-68.

[70] M. H. Wang and Z. Y. Lee, "Application of extension method and chaos theory in ECG identity recognition system," in *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 2016, pp. 1247-1251.

[71] H.-T. Yau and M. H. Wang, "Chaotic eye-based fault forecasting method for wind power systems," *IET Renewable Power Generation*, vol. 9, no. 6, pp. 593-599Available: http://digital-library.theiet.org/content/journals/10.1049/iet-rpg.2014.0269

[72] C. Chen, C. Lin, S. Lin, Y. Chiu, and C. Chiang, "A Chaotic Theorectical Approach to ECG-Based Identity Recognition [Application Notes]," *Computational Intelligence Magazine, IEEE,* vol. 9, no. 1, pp. 53-63, 2014.

[73] D.-F. Ge, B.-P. Hou, and X.-J. Xiang, "Study of Feature Extraction Based on Autoregressive Modeling in EGG Automatic Diagnosis," *Acta Automatica Sinica,* vol. 33, no. 5, pp. 462-466, 2007/05/01/ 2007.

[74] Z. g. Zhang, H. z. Jiang, D. f. Ge, and X. j. Xiang, "Pattern recognition of cardiac arrhythmias using scalar autoregressive modeling," *Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788),* vol. Vol.6, pp. pp. 5545-5548, 2004.

[75] O. Abdelhaq, E. Benachir, H. Aissaoui H. Aissaoui, and B. Bouikhalene B. Bouikhalene, *AR Modeling for Cardiac Arrhythmia Classification using MLP Neural Networks*. 2012, pp. 44-51.

[76] A. Ouelli, B. Elhadadi, and B. Bouikhalene, "Multivariate autoregressive modeling for cardiac arrhythmia classification using multilayer perceptron neural networks," *2014 International Conference on Multimedia Computing and Systems (ICMCS), Marrakech,* pp. pp. 402-406, 2014.

[77] F. Sufi, I. Khalil, and J. Hu, *ECG-based authentication*. 2010, pp. 309-331.

[78] S. Chauhan, A. S. Arora, and A. Kaul, "A survey of emerging biometric modalities," *Procedia Computer Science,* vol. 2, no. Supplement C, pp. 213-218, 2010/01/01/ 2010.

[79] B. Nasri, M. Guennoun, and K. El-Khatib, "Using ECG as a measure in biometric identification

systems," *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH), Toronto, ON,* pp. pp. 28-33, 2009.

[80] M. N. Dar, M. U. Akram, A. Usman, and S. A. Khan, "ECG biometric identification for general population using multiresolution analysis of DWT based features," *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec), Cape Town,* pp. 5-10, 2015.

[81] T.-W. D. Shen, W. J. Tompkins, and Y. H. Hu, "Implementation of a one-lead ECG human identification system on a normal population," *Journal of Engineering and Computer Innovations,* vol. 2, no. 1, pp. 12-21, January 2011 2011.

[82] N. V and S. Jayaraman, "Human Electrocardiogram for Biometrics Using DTW and FLDA," in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 3838-3841.

[83] N. Ghofrani and R. Bostani, "Reliable features for an ECG-based biometric system," *Proc. 2010 17th Iranian Conf. Biomedical Engineering (ICBME), Nov. 2010,* pp. 1-5, 2010.

[84] F. Agrafioti and D. Hatzinakos, "ECG based recognition using second order statistics," *Proc. Communication Networks and Services Research Conf. 2008, Nova Scotia, Canada,* 2008.

[85] H. Silva, H. Gamboa, and A. Fred, "One lead ECG based personal identification with feature subspace ensembles," *Proc. 5th Int. Conf. Machine Learning and Data Mining in Pattern Recognition, Leipzig, Germany,* pp. 770-783, 2007.

[86] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee, "ECG biometric recognition without fiducial detection," *Proc. Biometric Symp. (BSYM), Baltimore, MD,* 2006.

[87] O. Boumbarov, Y. Velchev, and S. Sokolov, "ECG personal identification in subspaces using radial basis neural networks," *Proc. IEEE Int. Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2009 (IDAACS 2009), Sept. 2009* pp. 446-451, 2009.

[88] Y.Wang and J. Yao, "A neural network to identify human subjects with electrocardiogram signals," *Proceedings of the World Congress on Engineering and Computer Science, San Francisco, CA,,* 2008.

[89] S. Saechia, J. Koseeyaporn, and P. Wardkein, "Human identification system based ECG signal," *Proc. TENCON 2005 IEEE Region 10,* pp. 1-4, 2005.

[90] R. Palaniappan and S. M. Krishnun, "Identifying individuals using ECG beats," *Proc. Int. Conf. Signal Processing and Communications 2004,* pp. 569-572, 2004.

[91] T.-W. Shen, W. J. Tompkins, and Y. H. Hu, "One-lead ECG for identity verification," *Proc. 2nd Joint Conf. IEEE Engineering in Medicine and Biology Society and the Biomedical Engineering Society, Houston, TX, 2002.,* 2002.

[92] A. Page, A. Kulkarni, and T. Mohsenin, "Utilizing deep neural nets for an embedded ECG-based biometric authentication system," in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2015, pp. 1-4.

[93] Q. Zhang, D. Zhou, and X. Zeng, "HeartID: A Multiresolution Convolutional Neural Network for ECG-Based Biometric Human Identification in Smart Health Applications," *IEEE Access,* vol. 5, pp. 11805-11816, 2017.

[94] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognition,* vol. 38, pp. 133-142, 2005.

[95] M. Kyoso, "A technique for avoiding false acceptance in ECG identification," *Proc. IEEE EMBS Asian-Pacific Conf. Biomedical Engineering,* pp. 190-191, 2003.

[96] J. Irvine, S. Israel, M. Wiederhold, and B. Wiederhold, "A new biometric: Human identification

from circulatory function," *Proc. Joint Statistical Meetings of the American Statistical Association, San Francisco, CA,* 2003.

[97] M. Kyoso and A. Uchiyama, "Development of an ECG identification system," *Proc. 23rd Ann. EMBS Int. Conf., Istanbul, Turkey,* 2001.

[98] J. Irvine *et al.*, "Heart rate variability: A new biometric for human identification," *roc. Int. Conf. Artificial Intelligence (ICAI' 2001), Las Vegas, NV,* 2001.

[99] C. Ye, M. Coimbra, and B. Kumar, "Investigation of human identification using two-lead electrocardiogram ECG signals," *Proc. 2010 4th IEEE Int. Conf. Biometrics: Theory Applications and Systems (BTAS), Sep. 2010,* pp. 1-8, 2010.

[100] R. Vullings, B. d. Vries, and J. W. M. Bergmans, "An Adaptive Kalman Filter for ECG Signal Enhancement," *IEEE Transactions on Biomedical Engineering,* vol. 58, no. 4, pp. 1094-1103, 2011.

[101] M. Chen, J. A. O'Sullivan, A. D. Kaplan, P.-H. Lai, E. J. Sirevaag, and J. W. Rohrbaugh, "Biometrics with physical exercise using laser doppler vibrometry measurements of the carotid pulse," in *2009 1st IEEE International Conference on Biometrics, Identity and Security, BIdS 2009, September 22, 2009 - September 23, 2009*, Tampa, FL, United states, 2009: IEEE Computer Society.

[102] S. Singla and A. Sharma, *ECG as Biometric in the Automated World*. 2010.

[103] P. R. Gomes, F. O. Soares, J. H. Correia, and C. S. Lima, "ECG Data-Acquisition and classification system by using wavelet-domain Hidden Markov Models," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 4670-4673.

[104] K. S, A. M, and S. M, *ECG Feature Extraction Techniques - A Survey Approach*. 2010.

[105] J. Yao and Y. Wan, "Improving computing efficiency of a wavelet method using ECG as a biometric modality," *International Journal of Computer Science and Network Security,* vol. 2, no. 1, pp. 15-20, 2010.

[106] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview Spectral Embedding," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 40, no. 6, pp. 1438-1446, 2010.

[107] F. Sufi, I. Khalil, and I. Habib, "Polynomial distance measurement for ECG based biometric authentication," *Security and Communication Networks Online,* vol. 3, no. 4, pp. 303-319, 2010.

[108] N. Ghofrani and R. Bostani, "Reliable features for an ECG-based biometric system," in *2010 17th Iranian Conference of Biomedical Engineering (ICBME)*, 2010, pp. 1-5.

[109] L. Ming and S. Narayanan, "Robust ECG Biometrics by Fusing Temporal and Cepstral Information," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 1326-1329.

[110] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee, "ECG Biometric Recognition Without Fiducial Detection," in *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, 2006, pp. 1-6.

[111] E. B. Mazomenos *et al.*, "A Low-Complexity ECG Feature Extraction Algorithm for Mobile Healthcare Applications," *IEEE Journal of Biomedical and Health Informatics,* vol. 17, no. 2, pp. 459-469, 2013.

[112] M. Niknazar, B. V. Vahdat, and S. R. Mousavi, "Detection of characteristic points of ECG using quadratic spline wavelet transfrom," in *Signals, Circuits and Systems (SCS), 2009 3rd International Conference on*, 2009, pp. 1-6.

[113] "Recommendations for measurement standards in quantitative electrocardiography. The CSE Working Party," (in eng), *Eur Heart J,* vol. 6, no. 10, pp. 815-25, Oct 1985.

[114] A. Chan, G. K. Isbister, C. M. Kirkpatrick, and S. B. Dufful, "Drug-induced QT prolongation and torsades de pointes: evaluation of a QT nomogram," (in eng), *Qjm,* vol. 100, no. 10, pp. 609-15, Oct 2007.

[115] A. A. Fossa *et al.*, "Dynamic Beat-to-Beat Modeling of the QT-RR Interval Relationship: Analysis of QT Prolongation during Alterations of Autonomic State versus Human Ether a-go-go-Related Gene Inhibition," *Journal of Pharmacology and Experimental Therapeutics,* vol. 312, no. 1, pp. 1-11, January 1, 2005 2005.

[116] S. G. Carruthers, B. McCall, B. A. Cordell, and R. Wu, "Relationships between heart rate and PR interval during physiological and pharmacological interventions," *British Journal of Clinical Pharmacology,* vol. 23, no. 3, pp. 259-265, 1987.

[117] J.-H. Atterhög and E. Loogna, "P-R interval in relation to heart rate during exercise and the influence of posture and autonomic tone," *Journal of Electrocardiology,* vol. 10, no. 4, pp. 331-336, // 1977.

[118] S. A. M. Said, R. Bloo, R. de Nooijer, and A. Slootweg, "Cardiac and non-cardiac causes of T-wave inversion in the precordial leads in adult subjects: A Dutch case series and review of the literature," *World Journal of Cardiology,* vol. 7, no. 2, pp. 86-100, 02/26

[119] M. Chen, J. A. O. Sullivan, A. D. Kaplan, P. H. Lai, E. J. Sirevaag, and J. W. Rohrbaugh, "Biometrics with physical exercise using Laser Doppler Vibrometry measurements of the carotid pulse," in *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*, 2009, pp. 1-6.

[120] H. Tenmoto, M. Kudo, and M. Shimbo, "MDL-based selection of the number of components in mixture models for pattern classification," in *Advances in Pattern Recognition: Joint IAPR International Workshops SSPR'98 and SPR'98 Sydney, Australia, August 11–13, 1998 Proceedings*, A. Amin, D. Dori, P. Pudil, and H. Freeman, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 831-836.

[121] Z. Liu, B. Malone, and C. Yuan, "Empirical evaluation of scoring functions for Bayesian network model selection," *BMC Bioinformatics,* vol. 13, no. Suppl 15, pp. S14-S14, 09/11 2012.

[122] L. L. C. Kasun, Y. Yang, G. B. Huang, and Z. Zhang, "Dimension Reduction With Extreme Learning Machine," *IEEE Transactions on Image Processing,* vol. 25, no. 8, pp. 3906-3918, 2016.

[123] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, *Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. 42(2), 513-529*. 2011, pp. 513-29.

[124] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, no. 1, pp. 489-501, 2006/12/01/ 2006.

[125] A. R. Hassan, "A comparative study of various classifiers for automated sleep apnea screening based on single-lead electrocardiogram," in *2015 International Conference on Electrical & Electronic Engineering (ICEEE)*, 2015, pp. 45-48.

[126] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine," *IEEE Transactions on Power Systems,* vol. 29, no. 3, pp. 1033-1044, 2014.

[127] M. Huanhuan and Z. Yue, "Classification of Electrocardiogram Signals with Deep Belief Networks," in *2014 IEEE 17th International Conference on Computational Science and*

Engineering, 2014, pp. 7-12.

[128] Y. Yan, X. Qin, Y. Wu, N. Zhang, J. Fan, and L. Wang, "A restricted Boltzmann machine based two-lead electrocardiography classification," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015, pp. 1-9.

[129] J. C. Sriram, M. Shin, T. Choudhury, and D. Kotz, "Activity-aware ECG-based patient authentication for remote health monitoring," presented at the Proceedings of the 2009 international conference on Multimodal interfaces, Cambridge, Massachusetts, USA, 2009.