HEART DATA ANALYSIS, MODELLING AND APPLICATION IN RISK ASSESSMENT

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF EDINBURGH NAPIER UNIVERSITY, FOR THE AWARD OF DOCTOR OF PHILOSOPHY IN THE FACULTY OF ENGINEERING, COMPUTING & CREATIVE INDUSTRIES

Owen Lo

May 2015

Acknowledgements

I would like to thank my Director of Studies, Professor William J Buchanan, and second supervisor, Professor Christoph Thuemmler. I would also like to thank all staff and researchers at Edinburgh Napier University who have provided the necessary support to enable this research to be completed.

Last but not least, I would like to thank my parents, Tak Ming and Shui Fong Lo, and my sister, Imelia Lo, for all the love and support you have given me.

Dedications

For S.M.K. In answer to your question, you know I do.

Declarations

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Contents

List of Figures

List of Tables

Abstract

The heart is a fundamental aspect of the human body. Significant work has been undertaken to better understand the characteristics and mechanisms of this organ in past research. Greater understanding of the heart not only provides advances in medicine but also enables practitioners to better assess the health risk of patients. This thesis approaches the study of the heart from a health informatics perspective. The questions posed in this thesis is whether research is capable of describing and modelling heart data from a statistical perspective, along with exploring techniques to improve the accuracy of clinical risk assessment algorithms that rely on this data.

The contributions of this thesis may be grouped into two main areas: statistical analysis, modelling and simulation of heart data; and improved risk assessment accuracy of the Early Warning Score (EWS) algorithm using a quartile-based technique. Statistical analysis of heart data, namely RR intervals, contributes to more informed understanding of the underlying characteristics of the heart and is achieved using null-hypothesis testing through the Anderson-Darling (AD) test statistic. The modelling process of heart data demonstrates methodologies for simulation of this data type, namely individual distribution modelling and normal mixture modelling, and contributes to assessing techniques that are most capable of modelling this type of data.

For improved accuracy on the EWS algorithms, a quartiles technique, inspired by anomaly-based intrusion detection systems, is presented which enables customisation of risk score thresholds for each patient defined dur-

ing a training phase. Simulated heart data is used to evaluate the standard EWS algorithm against the quartile-based approach. The defined metric of accuracy ratio provides quantitative evidence on the accuracy of the standard EWS algorithm in comparison with the proposed quartile-based technique.

Statistical analysis in this thesis demonstrates that samples of heart data can be described using normal, Weibull, logistic and gamma distribution within the scope of two minute data samples. When there is strong evidence to suggest that RR intervals analysed fits a particular distribution, individual modelling technique is the ideal candidate whilst normal mixture modelling is better suited for long-term modelling, i.e. greater than two minutes of heart data. In comparative evaluation of the standard EWS algorithm and the quartile-based technique using modelled heart data, greater accuracy is demonstrated in the quartiles-based technique for patients whose heart rate is healthy, but outside the normal ranges of the general population.

Acronyms

AD	Anderson-Darling
ANN	Artificial Neural Network
BPM	Beats Per Minute
CDF	Cumulative Distribution Function
CDSS	Clinical Decision Support System
CES	Continuous-Event Based Simulation
DARPA	Defense Advanced Research Projects Agency
DES	Discrete-Event Based Simulation
ECG	electrocardiogram
EM	Expectation Maximization
EWS	Early Warning Score
HPS	Human Patient Simulator
HRV	heart rate variation
IDS	Intrusion Detection System
IHR	instantaneous heart rate
IBI	inter-beat-interval

KS Kolmogorov-Smirnov

LARIAT Lincoln Adaptable Realtime Information Assurance Testbed

MEWS Modified Early Warning Score

MLE Maximum Likelihood Estimate

MOM Method of Moment

NEWS National Early Warning Score

PDF Probability Density Function

RTT Round-Trip Time

SW Shapiro-Wilk

TT Track and Trigger

VPS Virtual Patient Simulator

WFDB Wave Form Database

Chapter 1

Introduction

WITH the growth of information technology much of our everyday activities are now conducted using electronic devices and computing software. There is no exception to this trend in the domain of healthcare. Patient data, traditionally recorded in a paper-based format, is now not only managed in electronic databases, but the use of modern communication infrastructures, i.e. the Internet, has enabled rapid sharing of such records on a worldwide scale. In modern times, the provision of health care too has changed. Clinical Decision Support Systems (CDSSs), knowledge-based and machine-learning software, is a prime example of the use of information technology in the provision of improved healthcare. Through rapid computational abilities in modern IT systems, analysis of patient data has greatly improved along with understanding on how it may correlate with a person's wellbeing.

The work of the thesis continues the trend of improvements towards health care through the analysis and modelling of heart data along with its application towards improved risk assessment of a category of CDSS known as the Early Warning Score (EWS) system. Using statistical techniques, it is demonstrated that RR intervals may be modelled using a group of statistical distributions. Simulation of heart rate, derived from RR intervals, is used to evaluate a novel EWS algorithm which is capable of customis-

ing risk thresholds for each individual patient. The proposed algorithm is inspired by Intrusion Detection Systems (IDSs) - a form of security software used to detect computer threats. Furthermore, the success of this algorithm is also defined based upon evaluation methodologies and metrics from research conducted in IDS. Quantitative results from experimentation demonstrates evidence to support the argument that an anomaly-based EWS system provides greater accuracy in risk assessment of heart rate in comparison with the standard approach. The overall findings in this thesis provide greater insight to the underlying distribution of RR intervals whilst improved accuracy of the anomaly-based EWS algorithm shows there is potential for this approach to be improved upon by medical experts and even applied in real-life clinical environments in the future.

1.1 Background

This thesis stems from prior work conducted on the Cloud4Health [1] project. The primary aim of the Cloud4Health project was to provide a robust and scalable platform for the secure storage of electronic patient data. Appendix A gives further details on this area of work. The Cloud4Health project provided background on the subject of e-Health along with recognition on the shortcomings in two areas of research: vital physiological simulation and clinical risk assessment techniques.

Although an insight has been given on the average values of vital sign data, research towards realistic modelling of such data has yet to be fully realised. The work of Fox et al. shows that the mean value of heart rate is between 75 to 82 Beats Per Minute (BPM) while Pesola et al. demonstrates that systolic blood pressure is 112 mmHg [2]. Studies carried out by Mackowiak et al. and Shoemaker show that mean temperature of the human body is 36.8 °C [3, 4] whilst oxygen levels of a human body are generally between 96-98% as described by ODriscoll [5]. Finally, both Sherwood and

Tortora et al. agree that the mean respiration rate is found to be 12 breaths per minute [6, 7]. The average values of a human bodies vital physiological sign data may formulate the starting point for successful modelling of such data however further facts must be known to achieve an accurate model including the underlying distribution and characteristics of each vital sign data in relation to a bodies state of being. Given the significant scope of work required for such a goal, the choice was made to focus on one specific aspect for this thesis: the analysis and modelling of heart data.

Complimentary to the goal of heart data analysis and modelling, potential was noted to improve the EWS algorithm, one of the most popular clinical risk assessment techniques used in real-life healthcare environments, via computational power. The standard EWS algorithm defines risk values in a patient's vital physiological sign based on static predefined thresholds. The decision-making process is conducted using a series of conjunctions, e.g. if heart rate is less than 40 raise an alarm, therefore the assumption in this technique is that values of normality are the same for each patient assessed. This thesis argues that an IDS inspired technique using quartiles statistics can achieve more accurate results when assessing the health risk of a patient. The goal of increased accuracy in the electronic EWS system runs in parallel with the analysis and modelling of heart data, the key form of data that may be used to test and validate the improved EWS algorithm.

1.2 Research Aim and Objectives

This thesis aims to analyse, model and simulate heart data to evaluate a novel quartile-based EWS algorithm, which enables customisation of risk thresholds for each individual patient, to demonstrate increased accuracy on the proposed approach taken for risk assessment. It has been identified that heart data, especially heart rate variation (HRV), has been analysed and modelled in research of others but disparity of results in regards to the

underlying distribution is found. In regards to the EWS system, it is found that the current approach taken to risk assessment produces a number of inaccuracies due to the static approach of the algorithm.

This thesis argues that no one single distribution is capable of describing and modelling all heart data. Instead, it is proposed that a select group of statistical distributions may be used to encapsulate heart data in the general population. Similarly, like heart data, it is argued that the existing EWS algorithms do not provide a viable results in regards to accuracy due to the wide number of variation in patient's vital physiological sign data. It is proposed that a customisation approach to risk assessment, which considers the normal starting values of patient's parameters which are monitored, produces better accuracy in comparison to a static predefined rule set. To achieve the overall aim of this thesis, the follow three objectives are given focus:

- The establishment of simulation and modelling techniques applicable in healthcare and computer based simulation. Review on existing simulation and modelling methods applicable to healthcare need to be conducted in order to not only identify current trends, but show the novelty in the method applied in this thesis. Furthermore, identification on existing analysis and modelling methods for heart data need to be reviewed along with assessment on the shortcomings of the existing techniques employed.

- The creation of an improved algorithm for risk assessment in the electronic EWS system. A key goal to this objective is in the identification of not only limitations of the existing EWS algorithms but also show its current application in the context of health care. The techniques employed by both CDSS and IDS may be both applicable to improvements towards the accuracy of the EWS system. Thus, it is justifiable that a thorough review on both CDSS and IDS techniques is applica-

ble in this thesis.

- The creation of metrics which define experimental success in both heart data modelling and improved accuracy in the EWS algorithm. Evaluation and validation results must be shown for both the modelling method applied to heart data along with techniques towards demonstrating improvements to the EWS system. Metrics of evaluation need to be addressed in the literature to establish a standard approach to presentation of findings.

1.3 Contributions

In achieving the aim of the thesis, the three main contributions made are:

- Statistical analysis and formal hypothesis testing demonstrates that four primary distributions may describe small samples of RR intervals: normal, logistic, Weibull and gamma. RR intervals are the fundamental values of heart data analysed and modelled in this thesis. Qualitative and quantitative validation shows that RR intervals, modelled using one of the four identified distributions, is statistically similar to real-life counterparts. The finding that no one single distribution is capable of describing or modelling each patient's heart data is built upon previous work in the analysis of heart data including [8, 9, 10, 11].

- Recontextualisation of methodologies in evaluation of IDSs, especially the Defense Advanced Research Projects Agency (DARPA) evaluation [12] show that it is possible to apply such techniques to the evaluation of the EWS system in a quantitative manner. Metrics for evaluation in IDS, as originally presented by [13, 14], are modified to produce an accuracy ratio capable of assessing the sensitivity of the EWS

algorithm. The accuracy ratio is a measure of false-positives raised when analysing normal healthy patient's heart data.

- Contribution towards improved accuracy in the EWS algorithm produced two novel risk assessment techniques: a quartile-based and hybrid approach. Based on the method defined by [15] and inspired by the work of [16], the quartile approach is an anomaly-based IDS technique which demonstrates a higher degree of accuracy when assessing the risk of healthy patient's heart data which is outside the normal ranges of the general population. A minimum accuracy ratio of 0.14 was observed for the standard EWS algorithm while the quartile-based approach produced a minimum accuracy ratio of 0.66. However, inaccuracies are still produced using the quartile-based technique during some experiments that involved patients with heart rate within the range of normality. Thus, a hybrid approach, integration of both knowledge-based and anomaly-based techniques, demonstrates a higher degree of accuracy in comparison with both the National Early Warning Score (NEWS) algorithm and the quartile approach with a minimum accuracy ratio of 0.92 in experiments conducted.

1.4 Thesis Structure

The thesis is structured as follows:

- **Chapter 1. Introduction** - The aim and objectives of this thesis are presented. Contributions made in the work conducted in this thesis is presented too.
- **Chapter 2. Theory** - The core subjects of computer simulation, probability statistics, clinical risk assessment systems and fundamentals on heart data is provided to give background on the topics of this thesis.

- **Chapter 3. Literature Review** - The literature review gives focus to two main themes: simulation and CDSS. The first goal of this chapter is to explore current work related to simulation and decision support in the context of healthcare. Specific methodologies for analysis and modelling of heart data is reviewed along with CDSS techniques. Review of evaluations, i.e. clinical trials, conducted on the EWS outlines the need for improved methods in risk assessment which reduces the sensitivity of the algorithm.

- **Chapter 4. Heart Data Analysis** - Distribution identification of real-life patient's heart data, in the context of RR intervals, is conducted via formal hypothesis testing. Results give evidence that four primary distributions may be used to describe this form of data in small samples (up to two minute durations).

- **Chapter 5. Heart Data Modelling** - Methodology of heart data modelling via individual distribution is presented along with comparison with another common technique found in the literature named normal mixture distribution. Quantitative comparison of the two modelling methods with real-life patient data is provided to show the advantages and limitations of each approach.

- **Chapter 6. Heart Data Risk Assessment** - A quartile and hybrid approach to the EWS algorithm, which enables customisation of risk scores for each patient, is presented. Modelled heart data derived from the previous chapters are used to evaluate the accuracy of the three approaches to risk assessment. Results show that there is great potential in the quartile and hybrid approach to reduction of sensitivity, i.e. false-positives, compared to the standard NEWS algorithm.

- **Chapter 7. Conclusion and Future Work** - A conclusion is drawn from work conducted in this thesis. Areas of future work are defined which also aim to address any limitations noted in the work

conducted.

1.5 Publications

There are two publications which which are related to the core work of this thesis:

- **O.C.W. Lo**, J.R. Graves, and W. J. Buchanan. Towards a framework for the generation of enhanced attack/background network traffic for evaluation of network-based intrusion detection systems. In European Conference on Information Warefare 2010, 2010.
- **O. Lo**, L. Fan, W. Buchanan, C. Thuemmler, and A. Lawson. Towards simulation of patient data for evaluation of e-health platform and services. In PGNET 2012, pages 160-165, 2012.

Additionally, publications which provide background and context to this thesis include:

- L. Fan, W. Buchanan, C. Thuemmler, **O. Lo**, A. Khedim, O. Uthmani, A. Lawson, and D. Bell. DACAR platform for ehealth services cloud. In 2011 IEEE International Conference on Cloud Computing (CLOUD), pages 219-226. IEEE, 2011.
- L. Fan, **O. Lo**, W. Buchanan, E. Ekonomou, C. Thuemmler, O. Uthmani, A. Lawson, T. Sharif, and C. Sheridan. SpoC: protecting patient privacy for e-health services in the cloud. In eTELEMED 2012, The Fourth International Conference on eHealth, Telemedicine, and Social Medicine, pages 98-104, 2012.
- **O. Lo**, L. Fan, W. Buchanan, and C. Thuemmler. Technical evaluation of an e-health platform. In IADIS E-Health 2012, pages 21-28, 2012.

- **O. Lo**, L. Fan, Buchanan W., and C. Thuemmler. Conducting performance evaluation of an e-health platform. Isaas P. Kommers P. Issa, T., editor, Information Systems and Technology for Organizations in a Networked, chapter 16, pages 295-315. Society IGI Global Publishing, 2013.

Chapter 2

Theory

2.1 Introduction

A cross discipline of various fields of research need to be understood in order to achieve the aim of this thesis - modelling of heart data and its application towards evaluating a customisable version of the electronic Early Warning Score (EWS) algorithm using anomaly detection techniques. The core subjects include computer simulation, probability statistics, clinical risk assessment systems and fundamental knowledge of heart data acquisition and analysis via electrocardiogram (ECG) recordings. To provide background knowledge for later parts of the thesis, this chapters set out to provide theoretical concepts behind each of the noted subject areas. Overview on the Track and Trigger (TT) system is first given - the formal category in which EWS algorithms fall under. Simulation and modelling techniques is discussed to show the current widely accepted methodologies which are applied in this field of study. Probability statistics follows with focus on description on the primary distributions which modelled heart data is based upon in this thesis. Finally, a return to a clinical context is made, with overview given on the concepts behind ECG recordings, heart data terminology and units of measurement.

2.2 Track and Trigger Systems

The primary goal of TT systems is to monitor patient's wellbeing and provide some form of trigger if a health risk is detected. One of most commonly used TT system in hospitals throughout the world is the EWS algorithm. The EWS was initially developed by Morgan et al. [17] and is a risk based scoring system used by nurses and other healthcare staff in rating a patient's health status, e.g. in Accident and Emergency departments. Like all TT systems, the risk of mortality in a patient is generally correlated with a higher risk score.

As shown in a literature review conducted originally by [18] many variations on the EWS algorithm exist including the Modified EWS [19], Aggregate-weighted track and trigger system (AWTTS) [20] and Multiple-parameter track and trigger system (MPTTS) [21]. A more recent approach is the National Early Warning Score (NEWS) algorithm that attempts to standardise the approach to clinical risk assessment within UK healthcare environments [22].

This thesis acknowledges variations in the EWS system, but in-depth analysis on the differences between each approach is outside the scope of this work. Instead, latter parts of the thesis give focus to the two most commonly referenced variations: NEWS and Modified Early Warning Score (MEWS) system both of which are still currently used in the hospital environment within the UK. Table 2.1 presents the scores related to each parameter based upon the NEWS system.

2.2.1 Risk Score Calculations

Despite numerous algorithms being presented, calculation of risk scores within this form of TT system is relatively straight forwards. At periodic intervals, one or more parameters of a patient are monitored. Some

Table 2.1: National Early Warning Score System

Parameter	3	2	1	0	1	2	3
Respiration Rate (BF)	≤8		9-11	12-20		21-24	≥25
Oxygen Saturation (%)	≤91	92-93	94-95	≥96			
Oxygen Supplement		Yes		No			
Temperature (°C)	≤35.0		35.1 - 36.0	36.1-38.0	38.1-39.0	≥39.1	
Systolic BP (mmHg)	≤90	91-100	101-110	111-219			
Heart Rate (BPM)	≤40		41-50	51-90	91-110	111-130	≥131
Level of Consciousness				A			V,P or U

Table 2.2: NEWS Clinical Risk Outcome

Clinical Risk	NEWS score
Low	0
	Aggregate 1-4
Medium	Individual Parameter Scoring 3
	Aggregate 5-6
High	Aggregate 7 or more

of the most commonly monitored parameters include the five primary vital physiological signs of a patient including heart rate, blood pressure, body temperature, respiratory rate and oxygen levels [19]. The standardised NEWS system propose for UK healthcare environments also considers the attributes of oxygen supplement and level of consciousness in a patient.

Depending on the parameters values observed a number ranging from 0 to 3 is assigned for each parameter and an aggregation of the result gives the patients risk score. A risk score of between 0 to 1 relates to low risk, 5 to 6 medium risk and greater than 7 high risk [22]. Table 2.2 presents the NEWS system's risk outcome based on the aggregated score.

Traditionally, the EWS system - and its variations - requires manual calculation of a patient's risk using paper-based charts (Figure 2.1). With advances in computing, automated systems are now also viable which take advantage of IT infrastructure to automate the monitoring and calculation of risk scores. An implementation of the algorithm which makes use of modern technology for calculation of risk scores is referred to as a electronic EWS in this thesis.

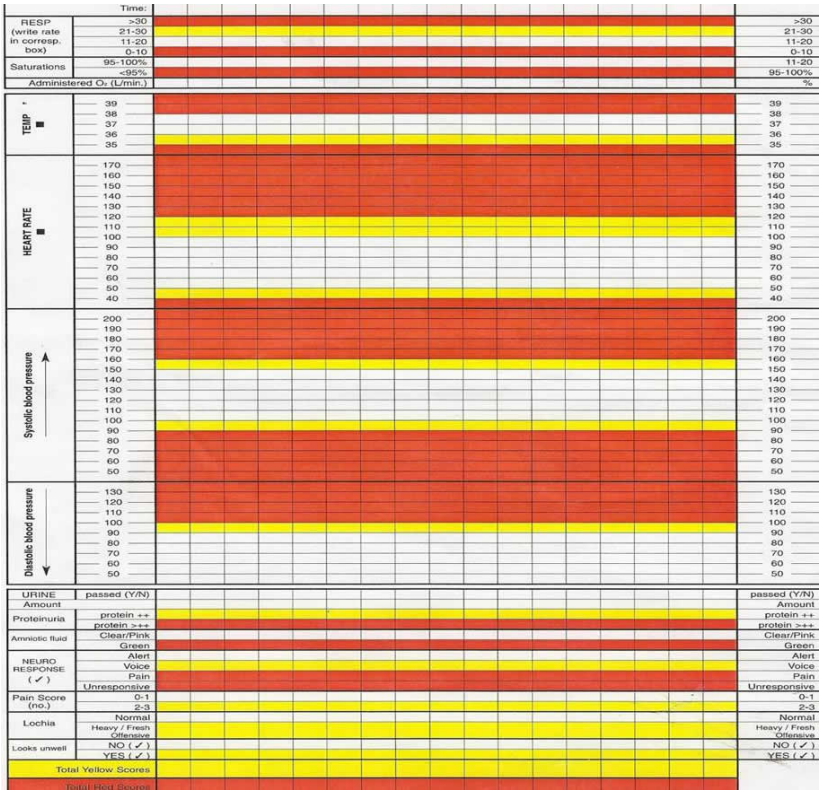

Figure 2.1: EWS Paper-based Chart

2.3 Simulation

Focus is given to fundamentals of simulation and modelling which are applicable to the work conducted on heart data in this thesis.

The concept of simulation dates back to many centuries with military establishments using simulation to conduct war games both for training and strategy decision making purposes. Perhaps one of the most well known traditional military simulations, which has evolved into a game still played today, is chess [23]. Similar to its roots, some of the first ever computer based simulation implemented were of a military nature [24] including the Manhattan Project [25] in which the simulation of nuclear detonations were carried out using the Monte Carlo algorithm [26].

Simulation techniques have progressed a significantly from its early beginnings. The applications of simulation is widespread, with many usages, including aerospace simulation [27], finance simulation [28] and, perhaps

most relevant to this thesis, medical simulation. Each method of simulation may differ based on the algorithms and implementation method used however, the end goal of each simulation will generally fall under one of the following categories: understanding a system, prediction of behaviour, training or entertainment. Furthermore, in some cases, simulation may breach more than one category, a prime example being Flight Simulators which can provide both entertainment and education for users.

2.3.1 Modelling

Ambiguities exist between the concept of simulation and the term modelling. In general discussions, it is common to find both terms used synonymously. However, although the differences are subtle, it is important to differentiate between a model and a simulation. This thesis shares the same views as Maria [29] in which a model is viewed as the representation of the inner workings of a system whilst the concept of simulation is the act or operation of the model for evaluation purposes.

A simple example of this in the context of healthcare is to consider computerised physical mannequins, otherwise known as Human Patient Simulators (HPS) used for medical training purposes. In this instance, the *model* is considered the physical representation of a human's body whilst *simulation* may consist of actions which the mannequin is capable of such as imitating breathing patterns which more advanced simulators are capable of. A prime example of a Human Patient Simulator (HPS) capable of such actions is the mannequins developed Laerdal [30] named SimMan® as depicted Figure 2.2.

2.3.2 Taxonomy of Simulation

In providing a taxonomy for computer based simulation, the work conducted by Sulistio et al. [31] group this subject matter into three main prop-

Figure 2.2: SimMan®: Human Patient Simulator [30]

erties: Presence of Time, Basis of Value and Behavior. Figure 2.3 presents the three main components of simulation.

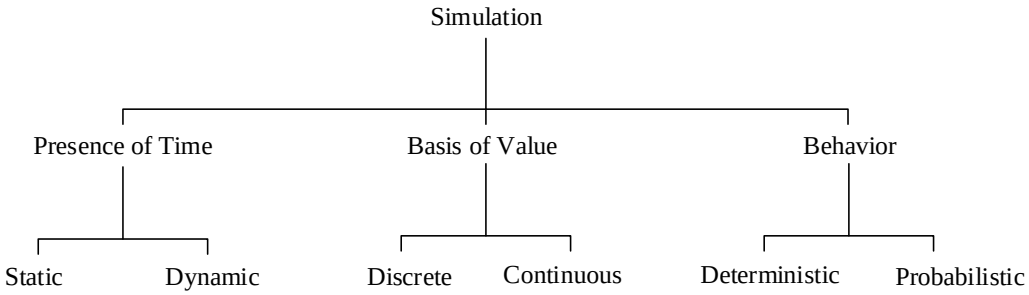

Figure 2.3: Simulation Core Components Taxonomy [31]

- **Presence of Time** relates to whether or not a simulation will consider the attribute of time. The concept behind the Presence of Time property is quite simple: either a simulation uses a time function (increasing or decreasing), or it does not. A simple example of presence of time is in the simulation of patient waiting times in a hospital environment. Such a simulation will consider presence of time important in order to keep track of how efficient the hospital environment is running.

- **Basis of Value** refers to the range of the entity modelled which the simulation is capable of generating. This can either be discrete enti-

ties or continuous entities. Discrete entities are limited in their range whilst continuous entities can have an infinite range. Simulating the growth of a human may consider the use of discrete entities, i.e. age in the unit of years, whilst simulating the act of running in a person may consider continuous entities since the body does not conduct movement at specific intervals.

- **Behavior** of simulation refers to whether or not repeated simulation using the exact same parameters, and values will result in the same results occurring. In other words, the behaviour of a simulation is either deterministic or probabilistic. In deterministic simulation, no random events will occur therefore, repeated simulations will always result in the same outcome. In probabilistic simulation, the opposite is true, and repeated simulations may produce slightly different results. Simulating the growth of cancer may consider probability based behaviour whilst simulation of a perfect hospital environment's waiting times may be more deterministic in nature.

2.3.3 Simulation Methods

The taxonomy of simulation in the previous section shows that three main components are required in the implementation of a simulator: presence of time, basis of value and behaviour. Presence of time is considered a binary choice, either a simulator uses a time keeping function or it does not. In the case of a presence of time being made available in a simulator, the basis of values are generated using one of two main techniques: discrete-event or continuous-event based simulation. Furthermore, regardless of time being a attribute of simulation, the method employed to simulate data, e.g. variables, will come under one of two methods: deterministic or probability based behaviour.

In applying a dynamic presence of time property, implementations of simu-

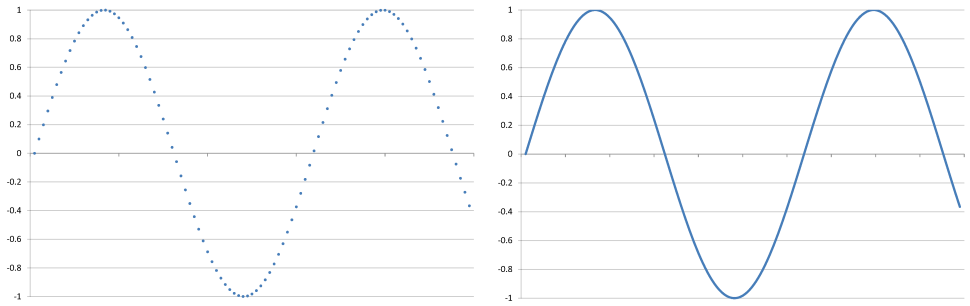

Figure 2.4: Discrete vs. Continuous Event Based Simulation

lations using discrete simulation is commonly referred to as Discrete-Event Based Simulation (DES) whilst continuous simulation is referred to as Continuous-Event Based Simulation (CES) [32]. The event refers to a simulation process taking place at a dynamic point in time which can either be specific time intervals, in the case of discrete simulation, or continuous time intervals, in the case of continuous simulation.

For example, consider a simulation which has a total duration of five seconds. Employing a DES method, an arbitrary timer interval of 1 second may be specified. Hence, at every 1 second interval, some form of processing, e.g. generation of a variable, in the simulation will be conducted for a total of five times, since the duration of simulation is 5 seconds. In comparison, employing a CES method will result in simulation processes being carried out continuously throughout the duration of simulation. Therefore, processing in the simulation begins from Time 0 and continues until a duration of 5 seconds has been completed. Figure 2.4 provides a comparison of these two techniques. On the left graph, a representation of discrete-simulation is shown where processing only occurs at specific time intervals whilst the right graph shows a continuous simulation whereby processing is conducted throughout the duration of simulation.

Since DES models variables which only change at specific defined points in time, it is well suited to simulating the events which take place at specific time intervals. CES will model events continuously, therefore this method

is better suited to simulating systems which will have constant changes [33, 34].

An example of a DES implementation can be found in the network simulator known as Ns-2 [35]. Ns-2 provides simulations of networks, including TCP, routing and multicast communication for research purposes. Ns-2 is defined as a DES [36] method since it only simulates network communication at specific intervals in time. In comparison, a stock market simulation using CES as the implementation basis is presented by Muchnik et al. [37]. Their justification for using CES is that stock prices will fluctuate in a continuous manner with no specific time intervals in which changes are conducted unlike the Ns-2 simulator. In other words, the stock market prices will rise or fall continuously throughout simulation.

Although simulation for a certain goal can be achieved using either DES or CES, the results of the experiments may differ drastically. As demonstrated in one study, both DES and CES methods were applied in simulating vehicle, i.e. a car, interactions by Jamison and McCartney [38]. In this study, it was found that widely different behaviors in the vehicles were produced depending in the basis of values. Using DES shows that the vehicle would behave in a chaotic manner, especially with increments in time interval whilst CES resulted in vehicles behaving in a more fluid manner. Hence, although neither method is better or worse, it can be stated that the choice of either DES or CES must be considered in great detail since the results may be skewered in a favourable or negative direction based on the method chosen.

In terms of calculation and output of results, the simulation can be considered either deterministic or probabilistic. It should be made clear that both DES and CES methods may be employed in producing deterministic or probabilistic calculations.

In deterministic behaviour, given one or more input parameters, calcula-

Figure 2.5: Deterministic Behaviour Example

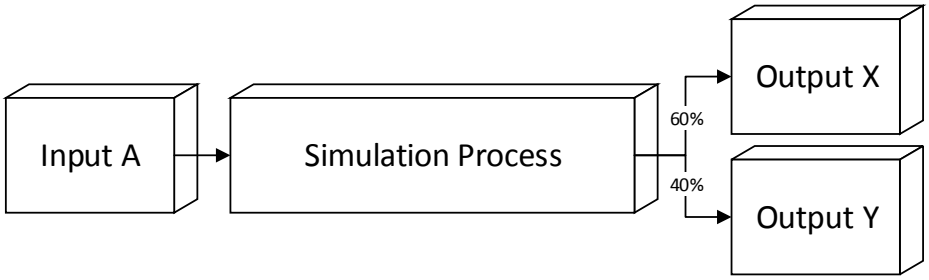

Figure 2.6: Probabilistic Behaviour Example

tion will result in either one, or a set number, of possible outcomes [39]. As an example, given the inputs a or b, input a will also result in output x whilst input b will always result in output y (Figure 2.5). Hence, repeatability and predictability of results [40] is the key attribute in deterministic simulation.

In comparison, the opposite is true of probabilistic behaviour. Probabilistic, also referred to as stochastic, behaviour introduces the concept of randomness in simulation [40]. Using a similar example as deterministic behaviour, given the input a, probabilistic results may return x with only 40% certainty or it may return result y with 60% certainty (Figure 2.6). Hence, there is a certain element of unpredictability in probabilistic behaviour.

Probabilistic behaviour is well suited to simulation where elements of uncertainty are required. As an example, a simulation of a slot machine (found in casinos) will require the use of probabilistic behaviour since there must be a degree of randomness in order to ensure users do not attempt to cheat the system. In comparison, a simulation of a factory production line

may employ deterministic behaviour since, based on the inputs of the simulation, it can be certain what the output produced by the production line may be.

In the case of implementing probabilistic behaviour, some form of random generation technique must be applied in order to provide the randomness of the simulation. One of the most well known techniques used is the Monte Carlo Method mentioned briefly in a previous section. In essence, this technique enables the of use repeated random sampling of numbers in order to come to a result [41]. It should be noted that there is no one singular Monte Carlo algorithm but instead, it is a group of mathematical algorithms.

2.3.4 Challenges in Successful Simulation

Analysis of the key challenges in conducting a successful simulation is given in this section. The work of Law [42] states that there are seven stages that must be carried out in the successful implementation of a simulator. The seven steps include formulation of the problem, gathering data, validating the model, programming the model, validating the implemented model, conducting and analysing experiments and presenting the results. Detailed description on each of these seven steps from the work of [42], supplemented with additional literature, follows:

1. **Formulation of Problem** - This first step involves defining what, exactly, the aim of the simulation is. One could argue that the formulation of a problem is not the challenge itself but rather it is defining how the simulation seeks to solve the problem. Hence, this first step is of great importance in providing a successful simulation since it provides the overall objectives of the simulation.

2. **Data Collection** - The second step to simulation is in conducting research to gather data that can be used as a baseline for the simulation.

In other words, this involves gathering information that the design and implementation of the simulator will be based on. It has been stated that data collection is perhaps one of the most time consuming aspects of conducting a simulation [43] with one source saying that it takes up to 40 percent of a project's time [44]. Though data collection is time-consuming, this step is fundamental as it is from the data gathered that a simulator will be modelled against. Thus, it is important to ensure the data collected is accurate since it will influence the simulators results [45].

3. **Validation of Conceptual Model** - From the data gathered, a theoretical model should be presented and validated to ensure it meets the aims of the simulation (as defined in step one). Validation, in this context, refers to ensuring that the right model is being built [46]. In other words, it is ensuring that the model presented is an accurate representation of the object(s) or system(s) which the simulator is simulating. Law formally refers to this step as conceptual-model validation [42] which is the process of ensuring any assumptions made about a model is accurate prior to implementation.

4. **Implementation of Model** - This step involves the actual creation of the simulation. The choice of either using a programming language or existing simulation software is entirely up to the developer's preference.

5. **Validation of Implementation** - Similar to step 3, upon successful implementation of the simulator, it is important to validate that the actual implementation is a valid representation of the model the simulator is attempting to simulate. The work of Law places special emphasis on two types of validation: face validity and results validation [42]. Face validity refers to whether the results *appear* to be correct from a subjective perspective while results validation assesses the im-

plementation in comparison with a real-life existing system [47]. In other words, validation can be conducted in two manners: qualitative validation and quantitative validation.

6. **Experiment and Analysis** - Once a valid implementation is achieved, experiments can then be conducted, and results analysed. The experiments carried out will relate back to what the initial aim of the simulation was (as stated in Step 1). After experimentation, the results can be analysed to determine whether the simulation was successful in solving the original problem.

7. **Presentation of Results** - The final and perhaps simplest step, assuming all previous steps have been carried out correctly, the results of the simulation can be documented and presented.

It is identified that the steps presented is similar to the iterative process conducted in a system development life cycles, e.g. analysis, design, testing, implementation, testing etc., and although carrying out simulation is a challenging task, breaking down the challenge into smaller steps makes the task much more achievable.

2.4 Probability Distribution

From the previous section, it has been shown that a wide variety of methodologies exist in the scope of simulation. The analysis, modelling and simulation of heart data conducted in this thesis makes significant use of continuous probability distributions thus this section provides some fundamental background to this subject. In particular, specific types of probability distribution including normal, logistic, Weibull and gamma distribution are covered. Figure 2.7 gives a visual representation on the four types of probability plots derived from the Probability Density Function (PDF) of each distribution. Definition of key terminology follows.

- **Cumulative Distribution Function (CDF)** - The Cumulative Distribution Function (CDF) describes the probability of a variable X occurring which is less than or equal to x within a given distribution. Each probability distribution will have a respective CDF. The CDF is defined as $F(x) = P(X \leq x)$ where F is the CDF of the chosen distribution. A CDF value should range from 0 to 1, i.e. 0% to 100% probability.

- **Probability Density Function (PDF)** - Whilst the CDF is capable of giving the probability of a variable occurring less than or equal to x, the PDF gives the density of a variable occurring in a given interval. The PDF, which is a derivative of the respective CDF, is used to calculate the probability of a variable X occurring that is within the range of $a < X < b$. Thus, if $f(x)$ represents the PDF, then the probability of a variable X occurring within a given interval $[a,b]$ can be represented as $P(a < X < b) = \int_a^b f(x)\,\mathrm{d}x$ so long as two primary conditions are met: 1) $f(x) \geq 0$ and 2) $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$ [48]. The term *density* describes how common a value is within a given distribution and it should be noted that this is a unit-less measurement. Inferring the probability of $P(a < X < b)$ from the PDF is given in Appendix B using normal distribution as example.

- **Parameters** - The term parameter relates to a quantifiable form of input which affects the output of a function. In the context of probability distribution, parameters will affect the overall form of the distribution including aspects such as shape, scale and location.

2.4.1 Normal Distribution

Alternatively known as Gaussian distribution [49], normal distribution is used to describe data which is both symmetrical and with data samples

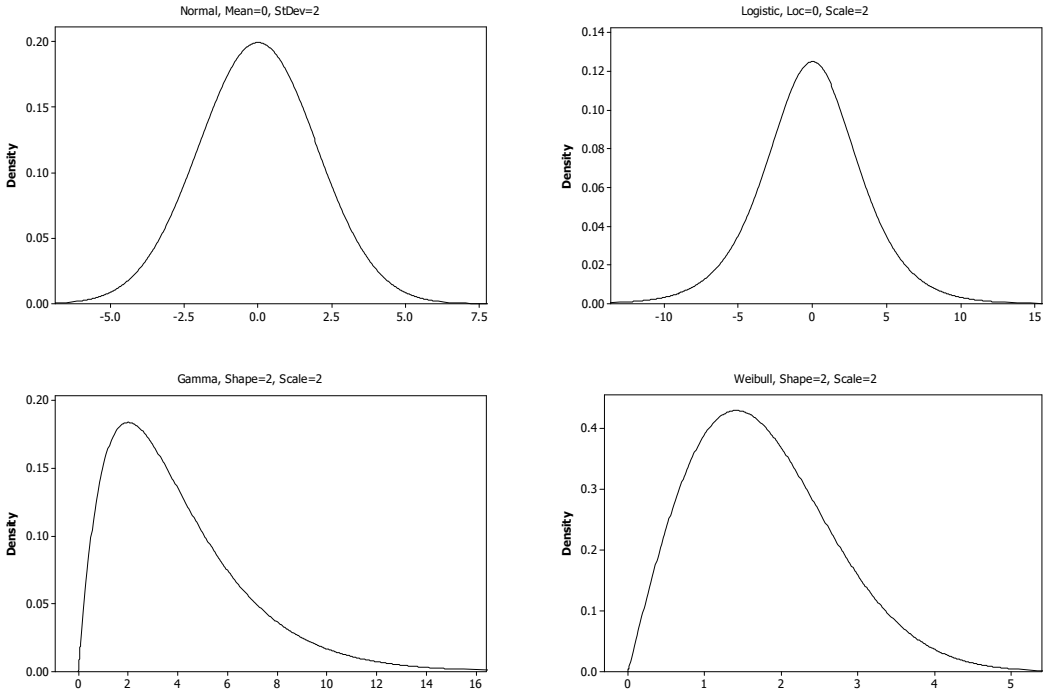

Figure 2.7: Probability Density Plot Examples

which cluster around a mean value [50, 51]. The two parameters for this distribution are μ (mean) and σ (standard deviation). Standard deviation, referred to as variance when squared, describes how wide spread the data is towards the mean value. The CDF of normal distribution is:

$$\frac{1}{2}[1 + erf(\frac{x-\mu}{\sqrt{2\sigma^2}})] \tag{2.1}$$

while the PDF is calculated as:

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.2}$$

2.4.2 Logistic Distribution

Logistic distribution is very similar in shape to normal distribution but with greater tails. In other words, wider variance will exist in logistic distribution in comparison with normal distribution. The two parameters of logistic distribution is s (location) and μ (scale). Location dictates the shift

in the distribution, relative to the x-axis, whilst scale dictates the pattern of the curve. The CDF of logistic distribution is:

$$\frac{1}{1+e^{-\frac{x-\mu}{s}}} \tag{2.3}$$

whilst the PDF is calculated as:

$$\frac{e^{-\frac{x-\mu}{s}}}{s\left(1+e^{-\frac{x-\mu}{s}}\right)^2} \tag{2.4}$$

2.4.3 Weibull Distribution

The Weibull distribution is of significant interest in this research due to previous application in modelling of heart rate variation (HRV) by [52]. The two parameters of Weibull are k (shape) and λ (scale). Unlike normal and logistic distribution, Weibull shares the same attribute with gamma in that the x-axis cannot be less than 0. It can be considered unique in that, unlike other distributions described in this section, there is no one specific shape or form which enables this distribution to be easily identified through visual analysis. In other words, it can be considered a general fitting distribution. The CDF of Weibull distribution is:

$$\begin{cases} 1-e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

whilst the PDF is calculated as:

$$\begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

2.4.4 Gamma Distribution

Gamma distribution tends to follow the characteristics of having a drop off in density of variables as the x-axis increases. This has resulted in its application in many economic type simulations to model stock markets and trades. The primary parameters of the gamma distribution are k (shape) and θ (scale) parameter. Depending on parameters, it can be often be seen that gamma looks visually very similar to Weibull. The prime difference in this distribution is that it has a slower or faster drop off point depending on the value of the k parameter. The CDF of gamma distribution is:

$$\frac{1}{\Gamma(k)}\gamma\left(k, \frac{x}{\theta}\right) \tag{2.5}$$

whilst the PDF is calculated as:

$$\frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}} \tag{2.6}$$

2.5 Heart Data Fundamentals

Monitoring and analysis of patient heart data is conducted using ECG monitors. An ECG monitor consists of various electro nodes which, when attached to specific parts of the human body, will pick up electrical impulses and report these results in a waveform drawn on ECG graph paper. The leads of an ECG monitor consists of, e.g. 3-lead, 5-lead or 12-lead, relates to the number of maximum electro nodes which the machine is capable of using. Figure 2.8 gives a visual representation on how the leads may be connected to the human body. Generally, having a greater number of leads will result in a more accurate waveform of a patient's electrical impulse. The electrical impulses of the human bodies heart activity is measured in amplitude (mV).

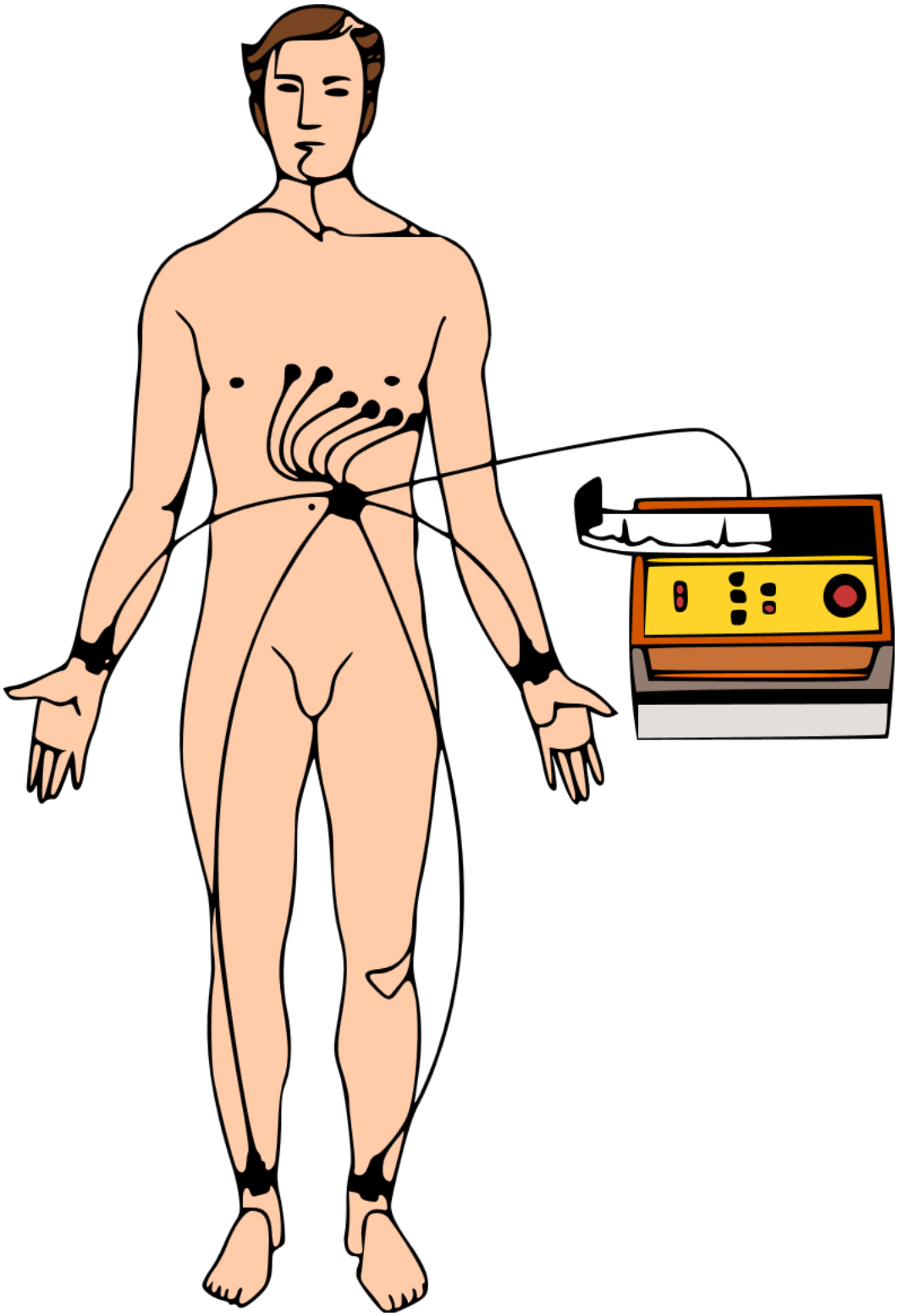

Figure 2.8: ECG attachments on the Human Body [53]

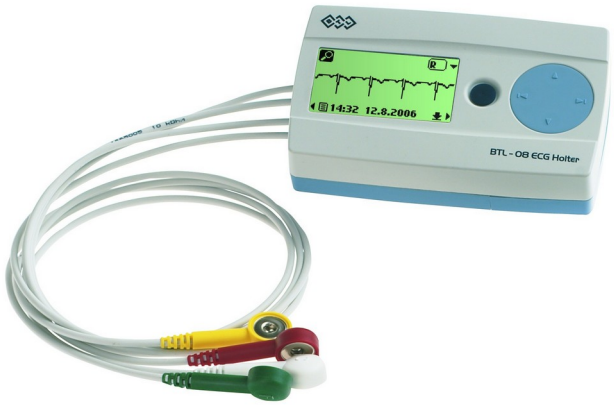

Figure 2.9: 4-Lead Holter Monitor [55]

Traditionally, heart activity of a patient is recorded on tracing paper - and still is - but advances in technology have enabled more mobile and compact versions of this monitoring device to be developed with one example being the series named Digital Holter Monitor [54]. Example of a portable Digital Holter Monitor is given in Figure 2.9. In the case of tracing paper recordings, the ECG signal is considered analogue whilst devices such as Holter will capture signals in a digital format. In the latter, a sampling frequency is defined which defines the level of capture detail (of electrical impulses) at each sample (of electrical impulses) per second. In other words, the representation of digital ECG data, in this thesis, is interpreted to be discrete values. Note that sample frequency is synonymous with the term *sample rate*.

2.5.1 ECG Waveforms and Terminology

The waveforms of a ECG follow a distinct pattern. One complete heartbeat of a patient can easily be identified based the unique pattern of a P-Wave, QRS complex, T-Wave and U-Wave as depicted in Figure 2.10. The

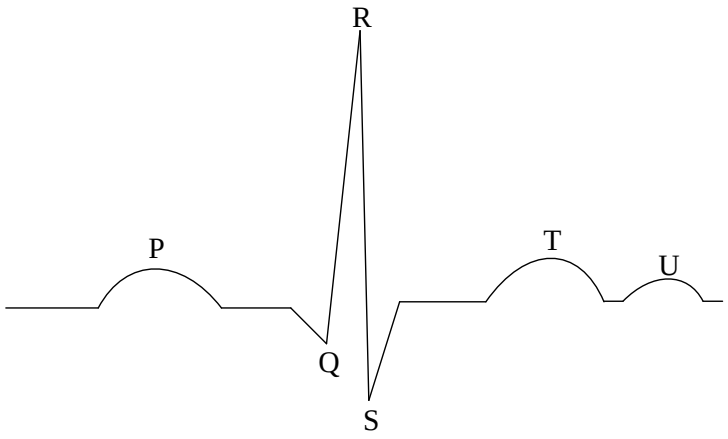

Figure 2.10: ECG waveform of one complete heart beat

unique peak of R, generally much sharper in comparison with the other waves, allows medical experts to easily identify each separate heart beat. The time taken between consecutive beats is referred to as the inter-beat-interval (IBI). When taken from ECG signals, this is more commonly referred to as the RR interval as shown in Figure 2.11. The RR interval is measured to be the time noted for a chosen R peak subtracted against its prior R peak. The RR interval is considered to be the one of the primary measures of HRV in a patient [56, 57, 58].

From the RR intervals one may derive the instantaneous heart rate (IHR) of a patient. The IHR is expressed as the average number of Beats Per Minute (BPM) - otherwise referred to as simply *heart rate* in this thesis. It can be stated that RR interval provide the fundamental base values for heart rate analysis in this thesis. The data source used for acquisition of ECG data along with conversion to different units of measurement is given next.

2.5.2 Data Acquisition and Unit Conversions

Description and methods applied here for acquisition and conversion of heart data is given with specific reference to digital ECG signals provided

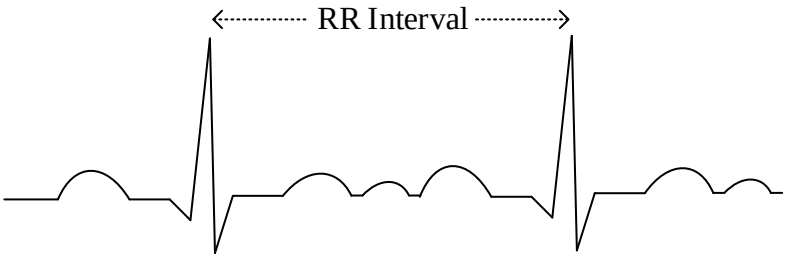

Figure 2.11: RR Interval in ECG waveforms

by Physionet [59] databank. This form of data is used during the experimental phases of this thesis for the goal of analysis and modelling of heart data. The purpose of this subsection is to give overview on Physionet's digital ECG data along with presentation of equations used in conversion of units; applied example of extrapolation of heart data from digital ECG signals using methods specified by Physionet is given in Chapter 4.

Physionet is a open-source database which contains a vast quantity of anonymised signals relating to the physiology of individual patients. The main data type of interest is ECG recordings in this thesis. The ECG records provided by Physionet are in a digital format. In other words, conversion from a continuous signal to a discrete signal has already been carried out. One of the key aspects of interpretation of the digital ECG records provided by Physionet is knowing the original sampling frequency (f_s) of the record. The f_s dictates the number of samples per second taken during the conversion of a continuous ECG signal to a discrete digital one. The optimal range for f_s is found to be between 250 Hz to 500 Hz [60, 61] for ECG recordings. Furthermore, in each of the records analysed for this thesis, an annotation file accompanies the digital record. The annotation file describes the time and occurrence of each individual heart beat in relation to the digital record. In other words, the R peak of each record is marked. The annotation file is created by medical experts who originally captured, analysed and uploaded the signal for public use.

The base unit of ECG recordings provided by Physionet is the RR sampling interval (i_s). The sampling interval relates to the time interval between each sample. In the case of ECG recordings, the i_s refers to the sample interval between two consecutive R peaks. To provide an example, assume that an ECG signal's f_s is 250 Hz. The standard equation for sampling frequency in relation to time T (in seconds) is as follows:

$$f_s = \frac{1}{T} \tag{2.7}$$

Therefore, sampling occurs every 0.004 seconds (1/250). If the i_s for two consecutive R peaks is known, e.g. 249, then calculation of $RR_{interval}$, expressed in units of seconds, can be achieved as follows:

$$RR_{interval} = \frac{i_s}{f_s} \tag{2.8}$$

Continuing with the example given, the result would be an $RR_{interval}$ of 0.996 seconds ($\frac{249}{250}$) which is the same as $i_s \times 0.004$. Once the $RR_{interval}$ is obtained, conversion to IHR can be conducted. The IHR is the average number of heart beats per minute in a patient. The unit of measurement of IHR is referred to as BPM and can be calculated as follows:

$$IHR = \frac{60}{RR_{interval}} \tag{2.9}$$

The IHR for this example would be equal to approximately 60.24 BPM ($\frac{60}{0.996}$).

Chapter 3

Literature Review

3.1 Introduction

THE literature review gives focus to the concepts of simulation and modelling, with emphasis given to related works within the context of healthcare. Techniques applied to analysis of heart data is reviewed, along with specific methodologies employed to model such data. The subject of Clinical Decision Support System (CDSS) becomes the focal point in the latter part of the review. The two primary implementations of CDSS, knowledge-based and machine-learning, is objectively compared. Within the technology of CDSS a focal point of this review is in analysing implementation's of the family of Track and Trigger (TT) systems more commonly referred to as the Early Warning Score (EWS) system. The final part of the review focuses on the concepts related to Intrusion Detection System (IDS) evaluation - a security software which is arguably similar to CDSS.

Review of simulation and modelling serves the goal of achieving the objective of heart data modelling whilst the concepts of CDSS serves to highlight the techniques which may be applicable to risk assessment. Review of the TT systems, in particular the EWS system, serves the goal of highlighting

the current issues related to this algorithm. Finally, review of IDS evaluation techniques is given since this technology forms the primary inspiration for the evaluation work conducted the latter chapter of this thesis.

3.2 Healthcare Based Simulations

A vast number of simulators are currently available, both commercially and for the purpose of research. It would not be conceivable to cover all simulators in existence, hence, in the scope of this thesis, a focus on simulation and modelling which are applicable in health care is given. In particular, this section of the chapter aims to present approaches simulators have been used to enhance healthcare environments.

3.2.1 Human Patient Simulators

Laerdal [30] presents SimMan®, a full-scale robotic mannequin, otherwise known as a Human Patient Simulator (HPS), which is capable of simulating the physical attributes of a patient. Dedicated software (Figure 3.1), which can either be run on a personal computer or a replicated patient monitor, allows for the simulation of vital physiological signs of the mannequin. Vital signs includes heart rate, blood pressure, temperature, respiratory rate and oxygen levels whilst the simulator is also capable of simulating the breathing patterns of a human being. In evaluations conducted on SimMan®, Hesselfeldt et al. [62] concluded that the simulated airway of the mannequin, though acceptable, was lacking in realism due to the mannequin having anatomic insufficiencies in comparison to a real-life human. In regards to the generation of vital signs, the studies carried out by Wyatt et al. [63] found that 36 out of 54 features of SimMan were rated at least average physiologically accurate by health professionals. Thus it can be stated that SimMan® has generally faced favourable reviews in the simulation of

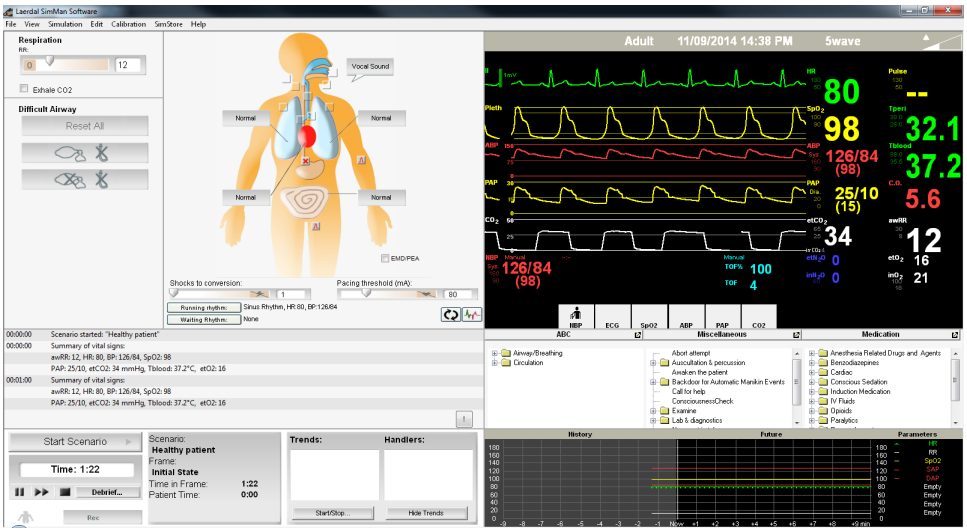

Figure 3.1: SimMan® Software [30]

a patient.

Hwang et al. [64] proposes the integration of both HPS and Virtual Patient Simulator (VPS) to provide a physical simulation of a patient, and a virtual simulation of the clinical environment. The HPS component of their work is a mannequin which models the clinical signs, e.g. heart rate, blood pressure and body temperature of a patient using scripts whilst the VPS component acts as a interactive clinical environment allowing users to control the mannequin with predefined commands. One novel feature of this work was the ability of the HPS to react to speech commands via the VP system. This feature is achieved using a *HPS Internal Data Exchange Protocol* (HIDEP), a Java based communication protocol, which enables two way communication between the virtual aspect and the physical aspect of the simulated patient. A high level depiction of the HIDEP protocols work flow is given in Figure 3.2.

Both the work of Laerdal and Hwang et al. present novel features in the implementation of simulated patients. However, both works conducted relate to the implementation of HPSs, whereby the main aim of such simulation is for the purpose of educating medical personnel for training purposes by providing a physical representation of a human body [65]. The work of

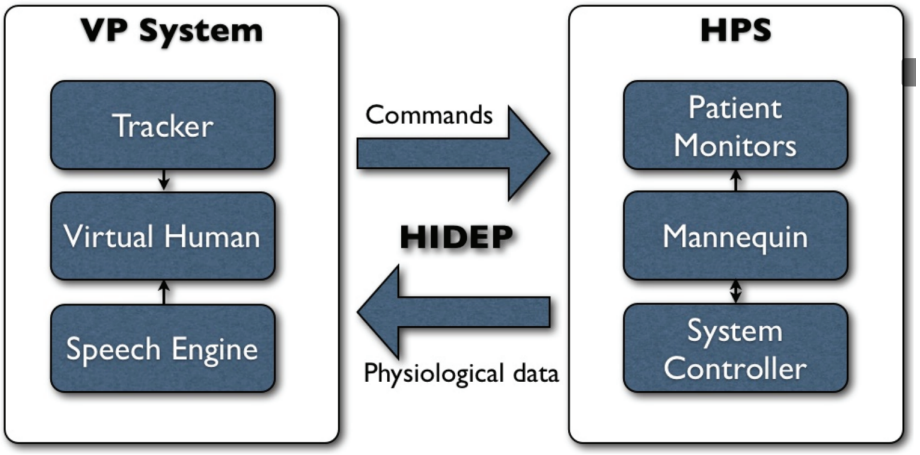

Figure 3.2: HPS and VP integration via HIDEP protocol [64]

Hwang et al. propose the use of VPs, but the purpose of this is for providing a script based approach in the control of their mannequin.

3.2.2 Traffic Flow Simulator

Using the technique of discrete-event based simulation, Meng et al. [66] propose the modelling of emergency hospital environments, i.e. Accident & Emergency, in order to understand traffic flow of patients and provide solutions to overcome overcrowding issues. Using a number of predefined scenarios, e.g. increment or decrement waiting times to see a consultant, and modelling patient arrival times using predefined schedules along with random arrivals, their simulation is capable of determining the time periods in which the increase in waiting times occurred. Example of factors which contributed to waiting times include scenarios such as patient consultancy time, number of available hospital beds and wait time for blood tests to be completed. In order to validate the simulation, simulated scenarios of both scheduled and stochastic arrival times where verified with medical staff's observations in their day-to-day work.

In similar work, Kolb et al. [67] propose five patient buffering concepts in order to reduce the traffic flow and overcrowding in emergency depart-

ments. Once again discrete-event based simulation was the technique chosen to test their five buffering techniques. The results of this work show show that the buffering techniques all improve upon the traffic flow of emergency departments. In terms of validation, the authors claim to follow the validation principles of work discussed previously including [32, 46] along with the work of Balci [68] discussed later in this review.

3.2.3 Vital Physiological Sign Simulation

With focus on work towards vital physiological sign simulation, Agar et al. [69] presents a simulation system which is capable of simulating blood glucose and insulin levels of both healthy patients and patients with Type 1 diabetes for educational and training purposes (Figure 3.3). Named GLUCOSIM, this simulations purpose is on the modelling of physiological compartments which relate specifically to blood glucose and insulin levels including within the heart, brain, liver and kidney. The key novelty in their work is that the simulator allows for the input of parameters including time of meals, carbohydrate intake, insulin dosage, patient's weight and so on. By running the simulation, correlation between input parameters and output of glucose and insulin concentrations in blood and liver can be observed. It is explicitly noted in this work that the tool is not designed for clinical decision making processes. Instead, it can be stated the primary aim in the work of Ager et al. is in providing better understanding of diabetes from the educational perspective.

An interesting note on GLUCOSIM is that mobile applications have succeeded this area in more recent work. As the demonstrated by Eng and Lee [70], approximitely 33% of mobile health apps achieve the same goal as GLUCOSIM within the iPhone store, i.e tracking blood pressure, carbohydrate intake and so on, while 22% are focued on provided education on diabetes. The primary difference between mobile health apps and GLU-

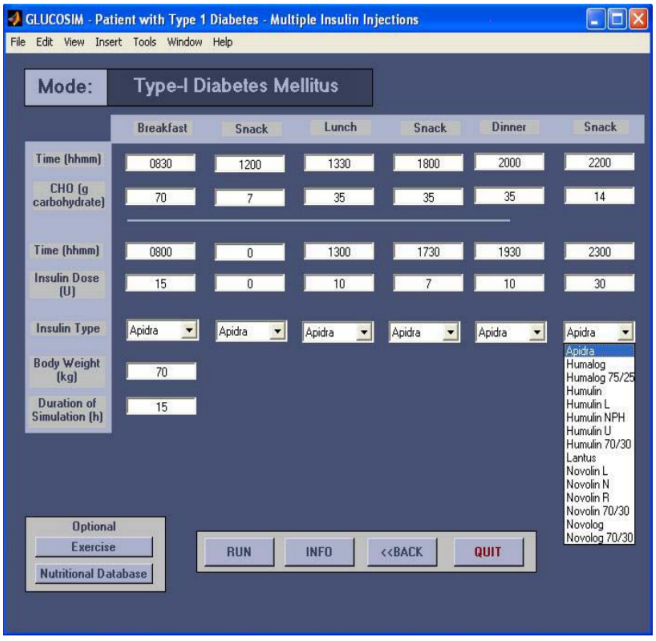

Figure 3.3: GLUCOSIM Software [69]

COSIM in that the input of data is achieved via user interaction rather than through the use of simulation.

3.3 Heart Data Analysis and Modelling

This section focuses on review of methodologies which have been applied specifically for analysis and modelling of heart data. It begins by reviewing the application of heart analysis before showing the existing work which has been conducted in analysis of heart data - which is achieved via statistical analysis of distributions. Formal methods of statistical analysis are reviewed as part of this process, with a focus given on specific techniques applied to modelling heart data in previous research.

3.3.1 Application of Heart Data Analysis

Analysis of heart data, especially RR intervals, by medical experts provides a wealth of knowledge about the human body, along with its overall well-

being. Research has shown that there is a direct correlation between respiratory rate and RR intervals by Brown et al. [71]. The impact of this study shows that the current physical state of a patient, e.g. standing, sitting, running, walking and so on, must be considered when analysing heart data since all these activities affect breathing patterns. This finding is further backed up by Tulppo et al. in which the authors show that RR interval times is related to the amount of physical exercise due to breathing patterns [72].

Analysis of RR interval also shows that there is a possible link between fractal-like, i.e. self-similar, patterns in the consecutive heart beats and the onset of atrial fibrillation (irregular and generally faster than normal heart rate) [73]. This infers that heart data may be a chaotic system. Better understanding between RR interval and possible heart issues has also lead to development of detection systems for abnormalities such as [74]. Other research on heart rate variation (HRV) has also shown many links to patient well-being, including higher risk of mortality after heart attacks [75] and a greater risk of infant mortality [76, 77] in when lower variation is observed. However, it must be noted the second finding has been disputed with findings by Antila et al. and Mehta et al. showing no significant correlation between infant mortality and variation of the heart [78, 79].

3.3.2 Statistic Analysis of Heart Distribution

The prior section has shown that potential correlation exists between heart behaviour and the detriment of a patient's health. Thus, analysis of heart data may provide significant information on the wellbeing of a patient. Additionally, to achieve the goal of successful modelling of RR intervals in this thesis, thorough understanding of its underlying statistical distribution is essential.

The study of irregularities in heart data conducted by Hashida et al. [9]

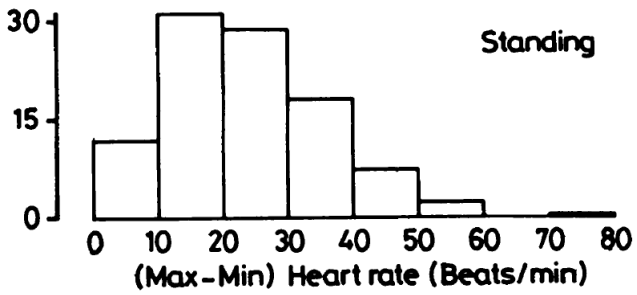

Figure 3.4: Positive Skewed Distribution of Heart Rate Variation [10]

shows that it is possible to classify a patient's RR interval with a gamma-like distribution during analysis of a 30 minute electrocardiogram (ECG) recording. In the work of O'Brien et al. [10], their analysis of 310 patient's HRV, over a period of 2 minutes, demonstrate a trend of positive skewered distribution, with a bias towards the right tail. Measurements included patient's being in a state of rest, inspiration (deep breathing), Valsalva (deep exhale), and standing. Figure 3.4 shows the result of positive skewness in heart data presented by O'Brien et al. when patients are in a state of standing. In a more recent study, the work of Mandrekar et al. [52] has hypothesised that RR intervals may be described statistically via the Weibull distribution.

Given the visual similarities between gamma and Weibull distributions, as shown in the Theory chapter, one may infer that either distribution type may be used to describe RR interval from the studies shown so far. However, a 24 hour analysis of RR interval by Tebbenjohanns et al. [11] showed that patient's with atrial fibrillation tend to follow a bimodal (two peak) distribution which differs entirely from both gamma and Weibull distribution. Additionally, in the work of Jennings et al. [8] their result shows that four patient's RR intervals may be described as normally distributed in 40-minute recordings in ten patients tested.

In regards to specific statistical methods, the primary technique that Hashida et al., O'Brien and Tebbenjohanns et al. used for concluding their findings

on heart distribution is through visual analysis of histograms. A more formal method, as applied by Jennings et al. and Mandreker et al., is the use of Kolmogorov-Smirnov (KS) test statistic. Given the qualitative nature of histograms, in comparison with the quantitative nature of test statistics, it can be stated that test statistics provide greater significance in proving the distribution of this data type while visual analysis compliments this goal by providing adding feedback of results in a visual manner.

3.3.3 Distribution Identification

It can be stated that heart data may follow many different forms of distribution. Identification of a statistical distribution in a set of data, e.g. RR intervals, is referred to as distribution fitting in this thesis. The concept of distribution fitting is concerned with calculating the probability that a set of data fits a certain statistical distribution. The primary technique for distribution fitting can be found in the work of Law [80] and Ricci [81], in which both authors agree that distribution fitting may be conducted in three primary steps, described as follows:

1. **Hypothesizing the Families of Distribution** - The first step in distribution identification is identifying the family of distributions in which the data may come from. This may be through visual analysis, e.g. plotting a histogram and observing the shape of the data, or through more quantitatively sound techniques such as calculating the underlying data's summary statistics including mean, median, variance, coefficient and skewness.

2. **Estimation of Parameters** - Having identified a family of distributions in which the a set of data best fits, the next step is to calculate the parameters of the distribution. Calculation of parameters serves the purpose of enabling one to generate random samples to compare

against a theoretical distribution, i.e. a goodness of fit test as described in the next step. Estimated parameter values also serve as the basis for modelling future heart data.

3. **Performing a Goodness of Fit Test** - A goodness of fit test is the formal technique used to hypothesize whether a sample of data comes from a specific distribution. It is a form of null-hypothesis testing where H^0 is the null-hypothesis and H^A is the alternative hypothesis. Many techniques may be used for this, whilst the two most relevant to this thesis include the KS and Anderson-Darling (AD) test statistic.

3.3.3.1 Parameter Estimates

Given a sample dataset where the precise distribution is unknown (but can be hypothesised as shown in the previous section), an estimate of the population parameter values based on the sample dataset can be achieved using parameter estimation techniques. Parameter estimates enable the generation of independent, and identically distributed, random variables for formal hypothesis testing via goodness of fit evaluation. Furthermore, estimated parameters allows for modelling of future samples of heart data as is shown in Chapter 5. Two of the most common applied parameter estimate techniques are Method of Moment (MOM) and Maximum Likelihood Estimate (MLE).

MOMs, first described by Karl Pearson in 1894 [82], is a technique for estimating distribution parameters based on comparing sample moments (of the data) with the population distribution moments. For k parameters in a population distribution, there must be $E(X_i^k)$ estimates which is a derived equation relating to the distribution's parameter at m_k moment. Generally, moments is the average power in the data samples of interest [83]. Thus, sample moments m_k of variables $\{x_1, x_2 ... x_i\}$ can be calculated as follows:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \tag{3.1}$$

As an example, assume variables $\{x_1, x_2 ... x_i\}$ follow a normal distribution with unknown parameters mean, μ, and variance, σ^2. MOM can be used to estimate the parameter where $E(X_i) = \mu$ and $E(X_i^2) = \mu^2 + \sigma^2$. Therefore, the derived equations for mean and variance are:

$$E(X_i) = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad E(X_i^2) = \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{3.2}$$

and solving them results in estimate for population mean, $\hat{\mu}$, and variance, $\hat{\sigma}^2$, as follows:

$$\hat{\mu} = \bar{X} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 (x_i - \bar{X})^2 \tag{3.3}$$

Another common parameter estimate technique is MLE. First proposed by C.F Guass in 1821 [84] before being reintroduced by R.A. Fishers in 1922 [85], ML estimates parameters by calculating, based on the sample data, values which has the most likely fit for a given distribution's probability density function [86]. The likelihood function may be represented as follows:

$$L(w|y) = f(y|w) \tag{3.4}$$

which states the likelihood of parameter w, given sample data y, is equal to function $f(y|w)$. The function to be applied is based on the probability density function of the distribution in which sample data y are observed to have originated from [86]. To put it another way, the likelihood of parameter w is based on a value of w which produces the highest probability of $L(w|y)$ occurring (hence *maximisation*).

In referring back to the variables $\{x_1, x_2 ... x_i\}$, which is assumed to follow a normal distribution, the unknown parameters mean and variance can be

calculated using the MLE by solving Equation 3.5. Note that the equation presents the Probability Density Function (PDF) of normal distribution. The example, and detailed proof in solving the equation which is out with the scope of this research, can be found in the work of Cassella and Berger (2002) [87]:

$$L(\mu,\sigma^2|x_1,x_2...x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{3.5}$$

Considered by some to be one of best parameter estimation technique [48], the MLE method is generally unbiased in larger sample sizes and has the smallest variance for parameter estimation [88]. This gives it the narrowest confidence interval, therefore the highest accuracy in estimates. In comparison, MOM has potential for strong bias, e.g. preference in calculating parameter values of sample data rather than population data. Furthermore, unlike MLE, the MOMs technique has the potential flaw of producing estimate parameters which are outside the bound of a distribution. An example of this may be seen in the work of [89] where method of moments was used to estimate parameter in log-normal distribution resulted in initial values which were negative which are generally meaningless in such a distribution - as log-normal does not accept negative values.

However, limitations also exist with the MLE method. One limitation to this technique is strong bias in small sample sizes, as demonstrated by [90]. At the same time, it is of note that they concluded that MLE was the best candidate for parameter estimates when sample size $n \geq 50$. A second limitation, in comparison with MOMs, is the non-trivial calculations which are necessary in the MLE estimate thus increased complexity and potential for mistakes to be made during estimation.

3.3.3.2 Goodness of Fit Test

The most commonly applied test statistic found in many areas of research is the KS test. The test statistic is:

$$D_n = \sup_x |F_n(x) - F(x)| \tag{3.6}$$

where $F(x)$ is the given Cumulative Distribution Function (CDF) to test for, i.e. if the null-hypothesis is normal distribution then $F(X)$ takes the form of this distribution's CDF. An alternative to the KS test is the Anderson-Darling test statistic. The AD test statistic, A^2, is calculated as follows:

$$A^2 = -n - S\,, \tag{3.7}$$

The variable S is derived from the following equation, where F is once again the CDF of the chosen distribution to test against, ln is the natural logarithm of F, and Y_i is a ordered list of the sample data:

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln(F(Y_i)) + \ln\left(1 - F(Y_{n+1-i})\right)\right] \tag{3.8}$$

Similar to KS, the AD test is used to define how likely a chosen sample of data fits a distribution by comparing the sample data's CDF against the chosen distribution's CDF [91]. In both cases, the tests are considered a null-hypothesis test, where H^0: *the sample data comes from the specified distribution* and the alternative hypothesis H^A:*the sample data does not come from the specified distribution*. Therefore, in the example of the AD test, if A^2 is greater than, or equal to, the critical value, the null-hypothesis is rejected and it may be stated that the data sample does not come from the distribution tested against. On the other hand, if A^2 is less than the critical value then there is not enough evidence to reject H^0.

In the context of distribution fitting, failure to reject H^0 does not give evidence that a set of sample data follows a specific distribution since the

alternative, H^A, is: *the sample data does not come from the specified distribution*. In other words, evidence has been produced to show that a sample of data does *not* follow a specified distribution. However, by testing multiple common distributions it is proposed that it can be shown which distribution has the highest likelihood of fitting. Critical values, for most of the common distributions, have been calculated in previous research by the author Stephens including [92, 93, 94, 95, 96, 97]. In practice, most statistical software which is capable of conducting AD test will already have critical values in-built thus removes the burden of looking up p-values in relation to critical values when conducting distribution testing.

Although both the KS and AD tests are capable of achieving the same goal - assuming the goodness of fit in a sample of data - one limitation subject to the KS test is the inaccuracy of results when there is significant outliers in the data [98]. Furthermore, research conducted by Razali and Wah [99] has shown that the AD test has far greater test power than the KS test in tests for normality in data. In the context of this thesis, the term *test power* relates to the probability of not accepting H^0 when it is false [100]. Failing to reject a false null-hypothesis is referred to as a Type II error [100]. Therefore, the evaluation conducted by Razali and Wah shows that the AD test has increased capabilities in correctly identifying a null-hypothesis and reduced probability in producing a Type II error. Finally, the work of Shapiro and Wilk [101] shows the AD test comes second only to the Shapiro-Wilk (SW) test but this test statistic is not ideal in most cases since it is designed to only test for normality in data samples [99].

3.3.4 Heart Data Modelling Techniques

With a focus specifically on the modelling aspects of heart data, three common techniques this thesis found include: mixture distribution, Weibull distribution and Artificial Neural Networks (ANNs) modelling. In mix-

ture distribution, one assumes that a distribution of data may be modelled by combining a group of distributions with different weightings. A common technique is to assume the entire distribution may be modelled via a normal distribution and perform a estimate of parameters for each curve in observed data. Such a technique is employed by both Costa et al. [102] and Ketchum et al. [58] for modelling of heart data. Costa et al. shows that it is possible to model heart data using three normal distribution components whilst the work of Ketchmum et al. proposes that such data can be modelled using only two components.

Alternatively, as shown in the work of Mandrekar et al. [52], there is the possibility of modelling heart data using only a single distribution: the Weibull distribution. The Weibull distribution is a flexible probability distribution capable of taking on many shapes. Mandreker et al. found that most heart data within the time frames of up to two minutes could be fitted to a Weibull distribution, leading to the hypothesis that Weibull distribution is capable of modelling small samples of heart data. Experimentation results in this work that marginal fits are achievable for most heart data modelled using the Weibull distribution.

A rather novel method applied to heart data modelling is in the use of ANN as demonstrated by Saalasti [103]. ANNs are based on the concept of synapses in the biological nervous system and consists of a series of nodes which are capable of learning patterns in data through various learning algorithms [104]. In essence, though more complicated algorithms exist, the simplest implementation of a ANN consists of various nodes which aim to fit data to a curve based on input data.

3.4 Clinical Decision Support Systems

The work of [105] defines CDSS as computer software which primarily aim to make decisions along with recommendations for a patient's clinical out-

come. Such results are directed at either the patient themselves or the clinician responsible for further action, e.g. proceed with a treatment plan or propose an alternative solution. The taxonomy of CDSS described by [106] found that 96% of such systems are developed for providing better patient outcome whilst only 4% of such systems are used for other purposes such as administrative tasks including calculation of clinical costs or resource utilisation.

Traditionally, much like paper-based patient records, CDSSs have been used in a standalone manner with no interoperability between different clinical environments. However, with the growth of e-Health technologies recent research including [107, 108] have proposed and shown the capability of sharing patient data between clinical environments and integrating CDSSs within differenc environments to enable rapid identification of patient illnesses.

The decision making methodology of CDSSs may be grouped into two primary categories: knowledge-based decision making and machine-learning decision making. In knowledge-based systems, experts in certain fields of medicine may input all their data into the decision support system and such system can then be used to decision making processes by other medical staff [109]. Alternatively, machine-learning algorithms in CDSS make use of artificial intelligence for the decision making process.

The work of Berner and La Lande [110] show that the most applicable technique for knowledge-based CDSS is rule-based implementations while machine-learning approach are generally applied using artificial intelligence techniques including ANNs and genetic algorithms. Figure 3.5 depicts the most common techniques applied to CDSS. Berner and La Lande's finding correlates with a study conducted by Berlin et al. [106] which the authors found that the most common implementation of CDSSs is rule-based techniques whilst ANNs and probabilistic models are also applicable. Given the wide scope of work conducted on CDSS implementation techniques,

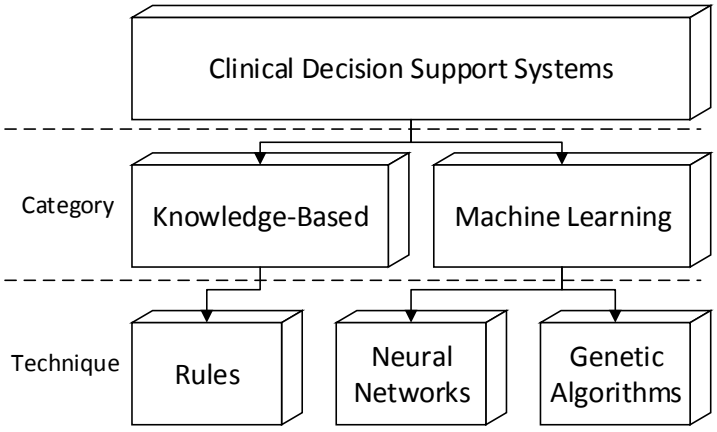

Figure 3.5: Categories and common techniques in CDSSs

this section focuses on review of the three most common approaches including: rule-based,ANN and genetic algorithms.

3.4.1 Rule-Based Approach

The most common knowledge-based approach to CDSS is the technique of a rule-based approach for decision making. One of the earliest examples of a rule-based technique using predefined expert medical knowledge, is the MYCIN project described by Buchanan et al. [111]. The primary goal of MYCIN is to produce an expert system capable of identifying and provided diagnostics on bacterial diseases. The name of this project is derived from the fact that many antibiotics will end with the suffix *-mycin*. MYCIN uses a set of semantic rules predefined as a series of *IF* and *ELSE* statements, thus leading to a logical flow during the decision making process. Expert knowledge on bacterial diseases was pre-populated in the system prior to the decision making process. A pseudo-code example of the rules implemented for MYCIN is:

```
IF: There is evidence that A and B are true
THEN: Conclude there is evidence that C is true.
```

Although the MYCIN project was never applied in live clinical settings, many researchers have followed similar approach in the implementation of CDSS. The work of Achour et al. [112] demonstrates the capabilities of rule-based CDSS by integrating it with the Unified Medical Language System (UMLS) with the aim of providing a standard approach to deciding what type of medical data should be shared between, and within, clinical environments. Their implementation is tested against the use case of deciding whether or not a patient requires a blood transfusion. Achour et al. evaluated their implementation by comparing their systems decision output against the decision a medical professional, i.e. a physician, would make in the same scenarios. More recent work by Rahaman and Hossain [113] show the author's using rule-based techniques for decision making on heart failure. Evaluation of their system shows that the CDSS produces less deviation in comparison with clinical history results in detection of heart failure.

One of the main limitations in the approach of rule-based implementations is the possibility of an incorrect decision made by the system. In one of the results presented by Achour et al., only 40% of physicians agreed with the CDSS decision hence showing the importance to ensuring accuracy in input data of both rules and expert knowledge during the implementation stage. Furthermore, as rule-based techniques make use of logical operations, i.e. IF and ELSE statements, as the complexity of the CDSS increases, so does the number of rules. In the MYCIN project by Buchanan et al., around 600 rules were implemented, whilst the heart failure detection system by Rahaman and Hossain made use of 216 rules. The work of Achour et al. does not provide a rule count, but given their goal of defining medical sharing decisions, it can easily be stated that the number of rules will grow significantly as the number of clinical environments are integrated with the system. Not only does maintenance of rules become difficult but it's proposed that rule-based systems may face the similar issue as computer

network firewall rules with anomalies such as overlapping, conflicting and redundant rules, as described by [114].

3.4.2 Neural Network Approach

As already discussed in Section 3.3.4, neural network is a form of artificial intelligence which makes use of numerous nodes for the learning pattern in data. In the context of CDSS, one example of the use of neural networks is in the detection of myocardial infarction in heart data. As originally described by Berner and La Lande, the neural network first goes through a training phase where it learns the patterns of normal and abnormal heart data. Testing and usage of the CDSS can then take place by feeding it new heart data. The decision on whether a patient displays symptoms of myocardial infarction will be decided based on the patterns recognised during the training phase. Naturally, such a technique is beneficial for early detection of any potential health problems with heart patients. Application of neural networks in CDSS is varied including over 300 articles of its application in cancer studies [115], determining well being in patients with diabetes [116] and even homeopathic medicinal systems [117].

One of the prime advantages of neural networks in comparison with rule-based system is the self-learning nature of this technique. This, in turn, eliminates the need for defining a set of rules thus eliminating potential issues in regards to human error. This, in turn, reduces the possibility of overlapping, conflicting and redundant rules [114] which rule-based approach may be susceptible to. Since training of neural networks is conducted through analysis of previous datasets, it also eliminates the need for predefined expert knowledge. Although one may see this point as advantageous it has been argued that the self-learning process eliminates accountability [110] in the decision making process if no (human) expert input is given along with difficulties in reverse engineering the logic behind

the neural networks decision due to complexity in the algorithms.

Furthermore, it has been argued that the complexity of neural network algorithms, from a medical perspective, may limit the use of this type of CDSS in healthcare environments [118]. Lisboa and Taktak demonstrates this fact in their review of neural network's application to cancer studies by stating that out of over the 300 articles reviewed only 27 studies resulted in clinical trials [115]. The overall impression on neural networks is that although numerous prototypes have been successfully implemented they do not fit into the work flow of existing healthcare environment thus the lack of adoption and usage in day-to-day clinical assessments [118].

3.4.3 Genetic Algorithm Approach

Genetic algorithms are derived from concepts of the biological organism and attempt to apply the theories of evolution and the principle of survival of the fittest [119] to computing-related problems. Through the process of reproduction and elimination using numerous methods, the theory of a genetic algorithm is that an optimal solution will be left which aims to address the issue it is attempting to solve.

Similar to neural networks, a wide number of research papers have been published in its application to CDSS including diagnostic of brain tumours [120] and diagnostic of heart diseases [121]. Furthermore, a data mining technique, through a combination of both genetic algorithms and neural networks, was applied by Amin et al. [122] for decision-making related to heart diseases in large a dataset of patient data. It can be stated that genetic algorithms share the same advantage as neural networks: there is no need for predefined expert knowledge nor the implementation of rules for the decision making process.

Unfortunately, genetic algorithms also share similar limitations as neural networks. Berner and La Lande backs this statement up by stating that al-

though *"genetic algorithms may be more accurate than the average clinician in diagnosing the targeted diseases, many physicians are hesitant to use these CDSS in their practice because the reasoning behind them is not transparent"* [110]. In other words, there is limited use of genetic algorithms in real-life healthcare environments due to the lack of understanding of the technique employed. Additionally, it can be inferred that the issue accountability, which affects neural networks, is also applicable to genetic algorithms due to their complex nature. Finally, referring back to the work Amin et al., it should be noted that although their experiments in the use of a genetic algorithm and neural network techniques produced positive results, no mention of live clinical trials or usage in real-life scenarios were noted.

3.5 Track and Trigger Systems

TT systems, especially the EWS algorithm, is one of the most common techniques used in healthcare environments for the assessment of a patient's health and wellbeing. Furthermore, the background section has shown that although such systems have been traditionally paper-based numerous research and development projects have been produced which have been successful in migrating the algorithm to electronic formats.

3.5.1 Application of the EWS System

Sufi et al. [123] proposes the use of mobile phone technology for automated calculation of the EWS algorithm, as employed by Walsall Hospitals NHS Trust. A standard client-server architecture is adopted whereby the client (mobile phone) acts as input medium for each parameter required by the EWS algorithm and the server (clinical infrastructure) receives the input via standard protocols including HTTP or SMS/MMS. The server calculates an aggregated score and returns clinical feedback to the client.

A similar methodology is employed by de Jager et al. [124] in which the authors propose the use of Wireless Sensor technology in place of mobile technology for automated calculation of the risk score. The primary difference in this work is the EWS algorithm used is the MEWS system rather than the traditional EWS system. Another similar field of work is conducted by O'Kane et al. [125] in which the authors propose the use of wireless body area networks (BAN) for automated capturing and monitoring of vital sign data using the MEWS system. A survey conducted on 71 health professionals showed that the system was widely more favourable in comparison with the traditional paper-based system, especially in regards to increased accuracy and maintenance of patient data along with reduction in paperwork and time saving.

3.5.2 Evaluation of EWS Systems

Each of the areas of work related to automated risk assessment in patient's using variations of the EWS algorithm are valid in their approach, however it does not address a fundamental issue related to the algorithm: effective and accurate assessment of patient risk and outcome. From existing literature, the work of Subbe et al. [19] shows that higher aggregated scores resulted in increased mortality rates in patient, but they also explicitly state that the Modified Early Warning Score (MEWS) algorithm should be analysed objectively by a clinical expert given the wide number of possible variation of results. A second study of the MEWS algorithm by Gardner-Thorpe et al. [126] concludes that the algorithm offered predictive capabilities in enabling rapid intervention for patients whose aggregated risk score were equal or greater than 4. Although the EWS algorithm applied differs slightly, similar results to the previous two research works were obtained in a study conducted by Goldhill et al. [127] where the authors show that increased abnormalities in vital parameters resulted in greater intervention

and mortality rates.

Despite the apparent positive effect of the EWS system for prediction of illness in patients conflicting opinions on this matter are found in more recent publications. In introducing a Medical Emergency Team (MET), one form of TT systems, the results of Hillman et al. [128] show that there was a greater number of overall emergency responses yet no correlation between incidences of mortality and abnormal vital sign parameters in the patient. Furthermore, a thorough review on all forms of TT algorithms conducted by Gao et al. [129] lead to the conclusion that not only was sensitivity in the algorithms poor, but no significant data exists to demonstrate strong evidence in the *"significance, validity, and utility"* of proposed EWS systems. In a similar review, conducted by Johnstone et al. [130], the authors also find that there is a lack of statistical techniques applied to defining risk assessment systems thus demonstrating a potential lack of rigour on defining what ranges of vital signs should be considered a risk along with providing a potential answer to the issue of sensitivity in the algorithm.

Finally, in a publication by Carberry et al. [131], the authors shares the same sentiment and agree with the view of Goa et al. that the EWS system, alone, is insufficient in providing accurate assessment on clinical outcomes due to factors including lack of communication between health care professionals and insufficient evidence to demonstrate efficacy when analysing patient's risk scores. Ultimately, the opinion of Carberry et al. is that the EWS system should only be used in conjunction with training professional's medical experience for increased accuracy of assessment.

3.6 Intrusion Detection Systems

In conducting review on CDSS and TT systems, it was noted that such systems share many similar characteristics with the computer security concept of IDSs. In particular, the experiments conducted on risk assessment

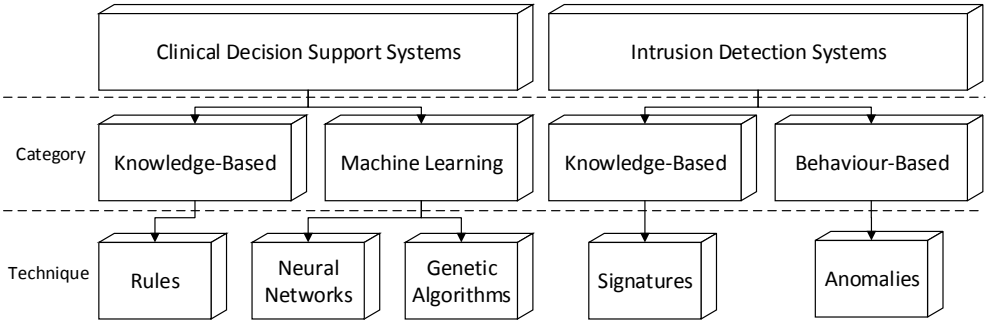

Figure 3.6: Comparison between CDSS and IDS

of heart rate in the later chapter, especially the metrics of evaluation, are greatly inspired by the methodologies applied to evaluation of intrusion detection systems. This section addresses similarities between IDS and CDSS along with review on methodologies which have been applied to evaluation of IDS.

3.6.1 Comparison to CDSS

In computing, IDSs are used for detection of threats and malicious traffic. In the taxonomy presented by Debar et al. [132], IDSs consist of two primary techniques for detection of threats behaviour-based and knowledge-based. Behaviour-based detection makes use of statistical analysis, and inference on detecting malicious traffic, whilst knowledge-based IDSs make use of predefined rules which contain signatures of known threats. As depicted in Figure 3.6, it can be stated these two detection categories of IDS are very similar to the concepts of knowledge-based and machine-learning found in CDSS.

Knowledge-based IDS, similar to rule-based CDSS techniques, consists of semantic rules, which are more commonly referred to as *signatures* in this field of research. In essence, signatures predefine known threats in computer systems. A prime example of a signature implementation comes from the IDS known as Snort [133] used for detection of malicious network traf-

fic. An example rule for Snort follows, which will simply produce an alert whenever TCP traffic is seen on the specified IP address:

```
alert tcp 192.168.1.1 any -> any any (msg:"Alert!";)
```

It can be stated the above rule is very similar to the concept of the electronic EWS system where an alert may be raised if a patient's heart rate is less than 40 Beats Per Minute (BPM) as would be in the case of the National Early Warning Score (NEWS) algorithm [22]. In comparison, anomaly detection techniques in IDS apply very much the same type of techniques as are applicable in machine-learning CDSS. From a study conducted by Garcia-Teodoro et al. [134] two common techniques, neural networks and genetic algorithms, are also applicable to behaviour-based IDS. Their work also covers a third technique in detection methods: statistical based approach. Literature review on neural networks and genetic algorithms has been conducted in prior sections of this chapter thus it is not repeated here for IDS since they share the same advantages and limitation. Instead focus is given to review on statistical approaches in the following section.

3.6.2 Statistical-Based Approach

One of the earliest works which proposes the use of statistical methods for detection of threats was proposed by Dennings [135] through the concept referred to as *Mean and Standard Deviation Model*. As the name implies, this technique involves the calculation of the mean and standard deviation of observed normal activity and for any future activity which falls outside this confidence interval may be considered anomalous therefore potentially a threat. Dennings also discusses another technique generalised as multivariate models which is based on profiling the correlation between two or more activities. By capturing the mean and standard deviation of all profiled activities, anomalous activity may be detected based on calculation of

confidence levels as before.

Following on from multivariate models, the work of Helman and Liepins [136] extends upon this idea through the concept of Bayesian analysis. In essence, Bayesian analysis makes use of statistical inference to calculate the probability of happening based on prior probability. In the case of detection of anomalous network transactions, the learning phase of the IDS will profile activity whilst the detection phase will apply Bayesian theorem to calculate the probability of the event occurring:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{3.9}$$

The primary advantage of Denning's approach of mean and standard deviation model is the simplicity of the technique. However, it does make the prior assumption that activity to be profiled follows a normal distribution which may not always be applicable, i.e. heart data is not always normally distributed. On the other hand, a Bayesian approach does not make assumptions on distribution of data and although the equation is easy enough to interpret, the primary challenge involves the correct calculation of prior probabilities which can be exceptionally difficult, even if sufficient prior data exists.

A more recent application of the statistical approach, as originally discussed in Wang et al. [16], is the use of quartiles and windowing techniques for detection of anomalies in data centres. Quartiles, originally formalised by Tukey [15], is the statistical technique of defining data range as 25% components, therefore thresholds can be obtained, whilst windowing involves segmenting the time series of data into collections of equal pieces. Thus, a combination of the two techniques enables one to identify outliers in data (thresholds) along with summary statistics for each windowed segment, e.g. distribution, mean, standard deviation, and so on. Quartiles, like the techniques of Helman and Liepins, make no assumption on dis-

tribution whilst the advantage of a windowing technique is that a data segment's distribution may be identified and compared with previously trained datasets for detection of any deviation of characteristics [16].

3.6.3 Evaluation Methodologies of IDS

A wide number of evaluation methodologies have been proposed for IDS. The two primary techniques for conducting evaluation is either real-time or offline evaluation [137, 13]. Real-time evaluation makes use of live network traffic and activities for assessing how well an IDS is capable of detecting threats, whilst offline evaluation makes use of datasets and simulation for evaluation purposes. Examples of real-time evaluation can be seen in the work of the Lincoln Adaptable Realtime Information Assurance Testbed (LARIAT) [138] and Trident framework [13], whilst work including the Defense Advanced Research Projects Agency (DARPA) evaluation [12] and research by Massicotte et al. [139], where a scripted approach is taken, make use of offline evaluation.

It can be easily stated that the prime advantage in real-time evaluation is in enabling realistic conditions during the testing process. Since threats are generated and applied to the IDS in a real-time manner, this methodology is better able to represent the actual environments in which an IDS may be deployed in [138, 140].

On the other hand, given the dynamic nature of live activities, reproduction of results may be difficult due to slight variations in the testing conditions. Reproducibility is perhaps the key advantage in offline evaluation methods. This is especially true of the DARPA evaluation, in which an offline dataset of both normal and malicious network traffic was made public for future research. Although the DARPA dataset has been critised due to being out of date [141], it is still one of the most widely cited and researched datasets within this field of work. Furthermore, it has been the inspiration

Table 3.1: Alarm Types in IDS

Alarm	Definition
True-Positive	Malicious activity has been taken place and alert has been raised.
False-Negative	Malicious activity has taken place but no alert has been raised.
False-Positive	No malicious activity has taken place but alert has been wrongly raised.
True-Negative	No malicious activity has taken place and no alarm has been raised.

for development of more recent offline datasets. A prime example is the KDD 99 dataset, produced by Mei-Ling et al., which claims a detection rate of 98% and less than 1% false-positive ratio during evaluation [142]. It is proposed that part of the reason for the popularity in offline datasets is due to the ease of availability (DARPA datasets: [143]; KDD 99 datasets: [144]) and the fact that there simply is no better alternative, as recently stated by Tavallaee et al.: *"due to the lack of better datasets, the majority of the research in the field of network intrusion detection is still based on the synthetic sets"* [145].

3.6.4 Metrics of Evaluation

Caswell et al. [146] defines the primary alarm types which an IDS is capable of triggering including true positives, false negatives, false positives and true negatives. Table 3.1 provides brief definition on each alarm type. Metrics of evaluation in IDS aims to describe how effective the implementation is in terms of each of these alarm types. Generally, an ideal result from evaluating an IDS should yield a high number of true-positives whilst false positives and false negatives should be limited.

The work of Sommers et al. [13], in presenting the Trident framework, proposes two primary forms of metrics for evaluation of IDS, the efficiency metric, defined as:

$$Efficiency = \frac{TruePositives}{AllAlarms} \tag{3.10}$$

and the effectiveness metric, defined as:

$$Effectiveness = \frac{TruePositives}{AllPositives} \tag{3.11}$$

Efficiency is a measurement on the overall performance of the IDS, i.e. false-positive ratio, whilst effectiveness is a measurement on the overall accuracy, i.e. false negatives, in the IDS [13]. The resulting value of 1 in either equation is ideal. Similar metrics are employed by related works including both the DARPA evaluation and LARIAT project where measurements consisted false alarms and attack detection rates [12, 138].

On the other hand, the work of Massicotte et al. make use of all four primary alarm types during evaluation but propose that the conjunction of these alarm types may allow for further classification of IDS metrics. A total of fifteen classifications are provided which may be split into three subcategories named success and failed attempts, failed attempts only and successful attempts only. Classification of detection metrics enables more meaningful and fine grained presentation of results in comparison with simply showing the number of true-positive or false-positive values as a numeric value.

Rab and Kalam [14] share a similar goal to Massicotte et al. when it comes to defining IDS metrics. In their work, they propose that detection metrics can be classified as both macroscopic and microscopic. Macroscopic metrics consist of detection ratio and false alarm ratio, calculated as:

$$DR = \frac{TruePositives}{AllPositives} \tag{3.12}$$

and:

$$FAR = \frac{FalsePositives}{AllAlarms} \tag{3.13}$$

whilst microscopic metrics relate to similar calculations but for each individual attack. It can be seen that detection ratio is the same metric as proposed by Sommers et al. thereby validating the correctness in this calculation whilst false alarm ratio can simply be derived from:

$$FAR = 1 - Efficiency \tag{3.14}$$

In other words, one can consider this the inverse of the efficiency metric.

3.7 Conclusions

The literature review has demonstrated that modelling and simulation, in a healthcare context, is primarily focused on the goal of providing of education and training. Similarly, it may be stated that this same goal is shared when applied to analysis and modelling of heart data. Review shows that modelling of heart data may be achieved via mixture distribution, Weibull distribution and neural networks from the works of [102, 58][52][103] respectively. On the analysis aspect of heart data, it is found that focus is given to analysis of HRV and a multitude of results are found with evidence to suggest that gamma [9, 10], normal [8], Weibull [52] and bimodal distributions [11] are all applicable.

Review of CDSS techniques shows that implementation of this system is achieved using one of two categories: knowledge-based and machine-learning. Review of one specific group of knowledge-based systems, TT systems, which encapsulates the electronic EWS show that the sensitivity of the algorithm is poor [129], there is a lack of statistical techniques in defining the potential risk of a patient [130] and not enough evidence exists to demonstrate a direct correlation between high risk score and further illness in a patient [128, 131]. The literature review demonstrates that CDSSs share many similarities with the computer security software known as IDS. In

particular, methodologies of clinical decision making and intrusion detection both apply the concepts of knowledge-based and machine-learning approaches. In the context of IDS, a commonly applied approach to detection of threats is through statistical analysis using anomaly detection techniques. One particularly relevant technique from this part of the review is the quartiles technique as applied by [16], which makes use of training datasets to define normality in data.

Given the many distributions found in heart data from the literature review, this thesis proposes that no one single distribution is capable of describing all heart data found in patients. The knowledge acquired on statistical analysis and modelling is brought forward in the next two chapters to demonstrate novel findings on plausible distributions which heart data fall under along with comparison of modelling techniques. In identifying the optimal techniques for modelling of heart data, instantaneous heart rate (IHR) values (derived from modelled RR intervals) are then applied to testing and validation of the NEWS algorithm. To address the issue of sensitivity in EWS algorithm as noted by [129], a novel approach to risk assessment is proposed which applies the knowledge-based technique of the EWS algorithm along with the anomaly detection technique of quartiles-based statistics as seen in the work [16]. The use of a statistical approach enables customisation of risk threshold for each individual patient's heart data analysed along with addressing the concern made by Johnstone et al. [130] that prior EWS algorithms do not consider the underlying statistics of vital physiological data when defining risk values. To provide for a benchmark comparison between the existing EWS algorithm and the novel technique, an accuracy ratio metric - inspired by IDS evaluation methodologies - is defined to provide a quantitative evaluation on each algorithm.

Chapter 4

Heart Data Analysis

4.1 Introduction

THIS chapter focuses on the analysis of real-life RR interval for identification of underlying distribution. RR intervals form the base values from which heart rate can then be derived from. The purpose of this objective to apply the knowledge obtained in order to model heart data for application in risk assessment. RR intervals are analysed in place of the derived instantaneous heart rate (IHR) values in order to give the most statistically accurate results in regards to the inter-beat-interval distribution of a human bodies heart. Since heart rate is simply the average beats per minute, it is argued that analysis of this data type does not give as much precision of underlying distribution in heart data in comparison with RR intervals.

Findings of the literature review have shown evidence of gamma [9, 10], Weibull [52] and normal distribution [8] in heart data. Based on this result, the primary hypothesis of this chapter is that one of four distributions are capable of describing the statistical distribution of RR intervals including normal, gamma, logistic and Weibull distribution. Distribution identification is performed using the using the Anderson-Darling (AD) test statistic,

in place of the more commonly technique referred to as the Kolmogorov-Smirnov (KS) test, due to being less susceptibility to outliers in data and demonstrating greater test power as shown in the literature review by [99] and [98] respectively.

4.2 Analysis Methodology

The methodology of Law [80] and Ricci [81], as discussed in the literature review, form the primary basis for the technique applied to distribution of heart data. Figure 4.1 depicts the methodology applied to heart data analysis:

1. Acquisition of real-life heart data.
2. Extrapolation of data to a format suitable for statistical analysis, i.e. RR intervals.
3. Hypothesis of distribution of acquired heart data, based on findings in the literature review.
4. Parameter estimation on the hypothesised distribution using Maximum Likelihood Estimate (MLE) method.
5. Goodness of fit, via the AD test statistic is applied to compare sampled values against real-life heart data samples.
6. Presentation of results is given in the form of AD value and related p-value to demonstrate whether there is significant evidence that the hypothesised distribution is plausible.

In reference to the last item, the success rate of the within the scope of experiments conducted in this chapter is defined in terms of whether there is significant evidence that one of the four hypothesised distribution is plausible in describing heart data analysed.

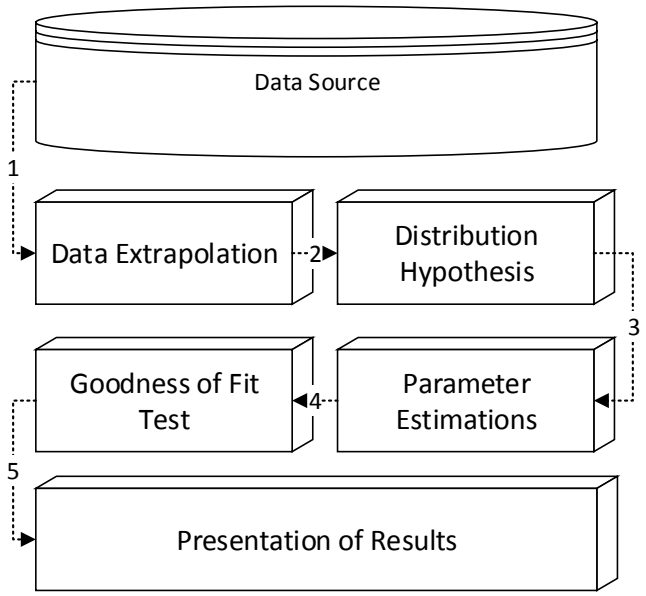

Figure 4.1: Steps in Heart Data Analysis

4.2.1 Data Source

It can be stated that analysis of patient's heart data provides significant indicators to a person's well-being [73, 74, 75]. In order to achieve successful modelling of heart data, for application within the Early Warning Score (EWS) risk assessment system, a key starting point is to acquire real-life data relating to RR intervals - the base value in which IHR may be derived from. The acquisition of such data enables the analysis of subtle behaviour in RR interval along with the ability to assess whether such data follows specific distributions.

In this thesis, acquisition of hearth data was achieved using the database known as Physionet [59]. Physionet is a open-source database which contains a vast quantity of anonymised signals relating to the vital physiological signs of individual patients. The data type of interest for this research is electrocardiogram (ECG) signals which allows one to derive the heart activity of a patient. The ECG signal of patients from the database, known as *Fantasia*, was acquired for the purpose of analysing RR interval data. The

Fantasia database was uploaded for public research purposes by Iyengar et al. [147] and their original work involved assessing age related differences in heart rate of both elderly and young patients [147]. Each record consists of approximately 120 minutes of ECG recordings, sampled at 250Hz, and all subjects remain in a resting state while watching the Disney movie Fantasia - hence the name of the database. In total, there are twenty young and twenty elderly patient records available in the dataset. To provide evidence of this chapter's hypothesis, that one of four distributions is capable of describing RR interval data, the scope of this thesis focuses on the first ten elderly (*f1o01...f1o10*) and young (*f1y01...f1y10*) patient's datasets. Therefore, a sample size of twenty datasets is used in this experiment.

To justify the choice of a sample size of twenty, it should be first noted that the work of Iyengar et al. makes a similar decision in their study of fractal dynamics of RR intervals where the sample size was twenty patients from the same database (i.e. ten elderly and ten young) [147]. In comparison, the work of Hashida et al. [9] analyses only one single patient while Jenning et al. [8] consists of a sample size of ten patients. More recently, the study of Mandrekar et al. [52] has a sample size of nine patients. Of the reviewed work on heart data analysis, only O'Brien et al. [10] has a greater sample size (310) of patients in comparison with this thesis - unfortunately this dataset is not made public. In this experiment, the use of twenty patients from the same publicly available database ensures not only reliability in results, since any other research may acquire the dataset and replicate the findings, but also ensures a level of control since all patients are deemed healthy by a medical expert and in the same (resting) state during acquisition of ECG data.

4.2.2 Data Extrapolation

As described in background of this thesis (Chapter 2), each record of the Fantasia database contains the digital ECG signal and an annotation file. Furthermore, a header file accompanies each record. The ECG signal itself is a digital representation of the patient's heart activity in the form of electrical impulses whilst the annotation file describes the time and occurrence of each individual heart beat in relation to the ECG signal. The header file provides meta data on the record. In viewing the header file for record *f1o01*, it can be seen that this record belongs to a 77 year old female, the duration of the signal is approximately two hours, and the sampling frequency (f_s) of the ECG signal is 250 Hz:

```
Record fantasia/f1o01
Notes
=====
 Age: 77 Sex: F
=====
Starting time: not specified
Length: 2:00:51.612 (1812903 sample intervals)
Sampling frequency: 250 Hz
```

Wave Form Database (WFDB) [148] is used to extrapolate the RR intervals from the ECG signal of each record, and designed specifically for analysis and visualisation of ECG signals. The software converts each ECG signal (stored in a binary format) into a human readable format using the *ann2rr* executable. Extrapolation of normal heart data from Patient f1o01 is given as follows:

```
ann2rr -r f1o01 -a ecg -c -P N -p N > f1o01.txt
```

The flag *-r* specifies the record to be read whilst *-a* specifies the annotation file related to the record. The annotation file consists of labels which point

to specific time of occurrence, and type, of each individual heart beat [149]. Annotations are manually added by the researchers, generally clinical experts, who chose to upload the record. The *-p* and *-P* flag enables filtering of specific heart beat types in the record. A value of *N* means normal heart beat only. The omission of the *-p* and *-P* flag will result in all beats being extrapolated, regardless of normality. A full list of annotation codes relating to types of non-normal heart beat may be found in the previous reference given. Finally, the *-c* flag is used to omit any RR intervals which are not annotated.

The results of using *ann*2*rr* to extrapolate normal heart data from Patient f1o01 results in 7166 data samples. The unit type for each sample is known as the RR sample interval (i_s). The RR sample interval is the interval between two consecutive R peaks. The i_s relates back to the initial f_s of the ECG signal which, as shown in Section 4.2.1, happens to be 250 Hz for each record in the Fantasia dataset. Thus, a i_s of exactly 250 means the RR interval for that sample is 1 second whilst a i_s of 243 would result in a 0.972 second RR interval. Conversion from i_s value to RR intervals, in the units of seconds, is achieved by performing the division of $\frac{i_s}{f_s}$ as described earlier in the Theory chapter of the thesis.

Table 4.1 shows the first five RR samples acquired using *ann*2*rr*, the equivalent RR interval for each sample and also the IHR expressed in Beats Per Minute (BPM) which can be derived from $\frac{60}{RR_{interval}}$. Of significant note in unit conversion is the possibility of rounding errors when converting IHR, in the units of BPM, back to RR intervals, in the units of seconds as originally discussed by Saalasti [103]. This would result in distortion of data and demonstrate the importance of ensuring the correct units are analysed. Table 4.2 demonstrates the issue of rounding errors which are applicable when converting IHR back to$RR_{interval}$ and $RR_{interval}$ to i_s. The same formulas as previously discussed are used in the manner where the reverse conversion of from IHR to $RR_{interval}$ is:

Table 4.1: Conversion of RR Samples at 250 Hz Sample Frequency

i_s	$RR_{interval}$	IHR
248	0.992	60.4
245	0.98	61.2
247	0.988	60.7
250	1	60
243	0.972	61.7

Table 4.2: Example of Rounding Errors in Reverse Conversions

IHR	$RR_{interval}$	i_s
60.4	0.9933775	248.344375
61.2	0.9803922	245.0981
60.7	0.9884679	247.117
60	1	250
61.7	0.9724473	243.1118

$$RR_{interval} = \frac{60}{IHR} \tag{4.1}$$

and the reverse conversion of $RR_{interval}$ to i_s is:

$$i_s = RR_{interval} \times f_s \tag{4.2}$$

In practice it is found that RR intervals are expressed most commonly in units of milliseconds but this chapter and other portions of the thesis presents findings and results of this data type in the units of seconds. Since the base unit for frequency is Hz, i.e. samples per *second*, conversion to milliseconds would need to be applied at some point in the equation to present the results in this manner. The decision was made not to perform this conversion to eliminate the possibility of distortion and irregularities which may occur as Table 4.2 has demonstrated. Furthermore, given that past medical studies on this type of data, examples including [150, 72, 147, 151], have

presented RR intervals in the units of seconds justification exists that this choice is not completely against standard convention.

4.2.3 Hypothesizing RR interval distribution

It can be made apparent, from the literature review, that attempting to classify heart data under one specific distribution is simply too naive an approach. There are many factors which result in different distribution of heart data in patients including age, gender, existing diseases and even the time period of the recording. Based on the findings of the literature review, it is hypothesised that RR interval data may be described as one of four primary distributions including normal, gamma, logistic and Weibull.

The choice of gamma and Weibull distribution are derived from results produced by Hashida et al. [9], O'Brien [10] and Mandrekar et al. [52] where it was found that heart variation tends to follow a skewered distribution. The choice of normal distribution is derived from the work of [8] where it is found that a normal-like distribution is demonstrated in heart rate variation (HRV) in four out of ten patient's data analysed in their work. Logistic distribution is also chosen since this distribution is very similar to normal distribution. It is argued that logistic fit will preside over a normal fit if the tails of the data analysed are longer - a highly likely case in instances where RR intervals have greater time differences.

4.2.4 Parameter Estimates

The choice of MLE for distribution identification when analysing RR intervals was made since analysis of every dataset in this thesis had a sample size $n \geq 50$, consequently mitigating the potential bias of MLE estimations originally described in the work of [90]. Furthermore, given that MLE has greater accuracy, in comparison with Method of Moment (MOM) [88], it

is natural to make use of this technique for parameter estimation. Finally, to address the complexity of calculations required in MLE, as discussed in the literature review, mitigation of the non-trivial calculations is overcome since almost all statistical software packages come inbuilt with the MLE algorithm. Thus, this thesis makes use of an existing implementation which also has the advantage of reducing the potential of any inaccuracies in calculation of results.

4.2.5 Goodness of Fit Test

The AD test statistic is applied for testing of distributions in RR intervals. This differs in the technique of previous work including [8, 52] since the KS test was used. From the literature review (Section 4.2.5), it has been shown that the AD test comes only second to the Shapiro-Wilk (SW) in test power [99]. The SW test is dismissed for use in this thesis since it is designed to be applied only under the assumption of normality in data.

Additionally, it has been shown that KS is subject to inaccuracies in distribution identification when there are outliers in the data [98]. The KS test, in comparison, does not share this same limitation. Given the nature of the data which the AD test statistic is to be applied against, i.e. RR intervals, where significant variance may exist in a patient's heart beats, it can justified that this test statistic is the most ideal for producing accurate results in the identification of the distribution.

4.3 Identification of normal healthy patients

For each of the ten young and elderly patients in the Fantasia dataset, distribution identification via the AD test for the full sample of each patient's RR interval data (approximately 2 hour recording) was conducted. All statistical analysis was conducted using Minitab 17 [152] and the significance

level for p-value is 0.05 for each test. The AD test statistic is automated generated by Minitab 17 when analysing data. To provide proof of validity in Minitab's calculation, Appendix C presents evidence that the AD test statistic, calculated manually, and the test statistic produced by Minitab are the same.

Distribution identification of patient's 2 hour RR interval proved inconclusive. A lack of fit for any of the hypothesised distributions i.e. normal, logistic, Weibull, gamma, was produced. As example, Figure 4.2 demonstrates the probability plot of Patient f1o01. The Y-axis represents the Cumulative Distribution Function (CDF) of each distribution tested while the X-axis represents the RR intervals values expressed in the unit of seconds. With a p-value of < 0.05 for each of the tested distributions, this result demonstrates that there is not enough evidence that the RR interval data comes from any of the tested distributions.

The possible reason for this result is due to the unpredictable nature of a patient's physiology. Even in patient's ECG recording, where they remain motionless, external stimuli such as sound and motion will result in changes to their resting heart rate over certain periods of time as shown in the work of [10]. Furthermore, general fitness levels [153] and even posture [154], will result in variation with heart rate over short periods of time.

4.3.1 Small Sample Identification

Due to this initial finding, the distribution identification method is refined by analysing smaller samples of the each patient's RR interval. The point at which RR interval distribution was successfully identified for most patients was found to be a maximum of 120 seconds. It can be stated that the methodology applied here, i.e. two minute sample analysis, is similar to the work of Mandrekar et al. [52] however the technique applied here differs, as the aim of this work is to assess whether RR intervals may fit

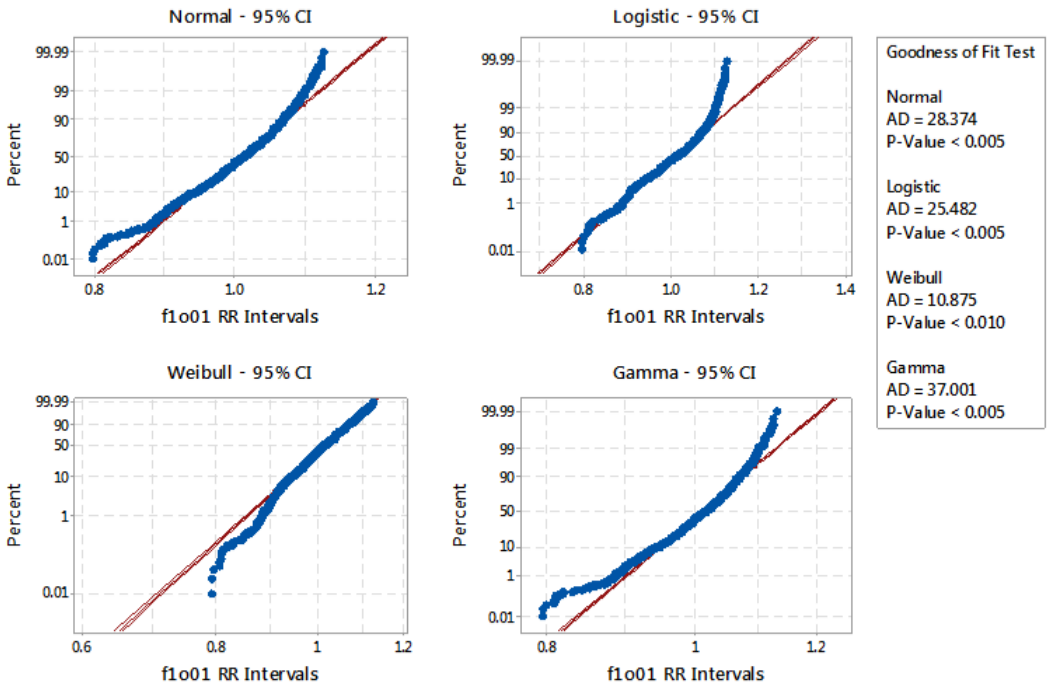

Figure 4.2: Two Hour Sample Distribution Identification of Patient f1o01

families of distribution rather than just Weibull distribution. Furthermore, the test statistic of AD is used rather than KS which, as the previous section explains, is considered to have greater test power in distribution identification.

Two minute samples of patient's RR interval data are acquired directly from the datasets by once again using the *ann2rr* command:

```
ann2rr -r f1o01 -a ecg -c -P N -p N -t 120 > f1o01.txt
```

where *-t* defines the time in seconds for the extrapolation process to stop at. Although this results in small differences in sample size n in each dataset to be analysed, the results enabled more accurate result of a patient's RR interval distribution over small periods of time. The small difference in sample size results from the fact that each patient has variations in the time taken for each individual heart beat. Thus a lower heart beat intervals will will result in less RR intervals within two minutes in comparison with a patient who has a higher heart beat interval. Figure 4.3 shows the results

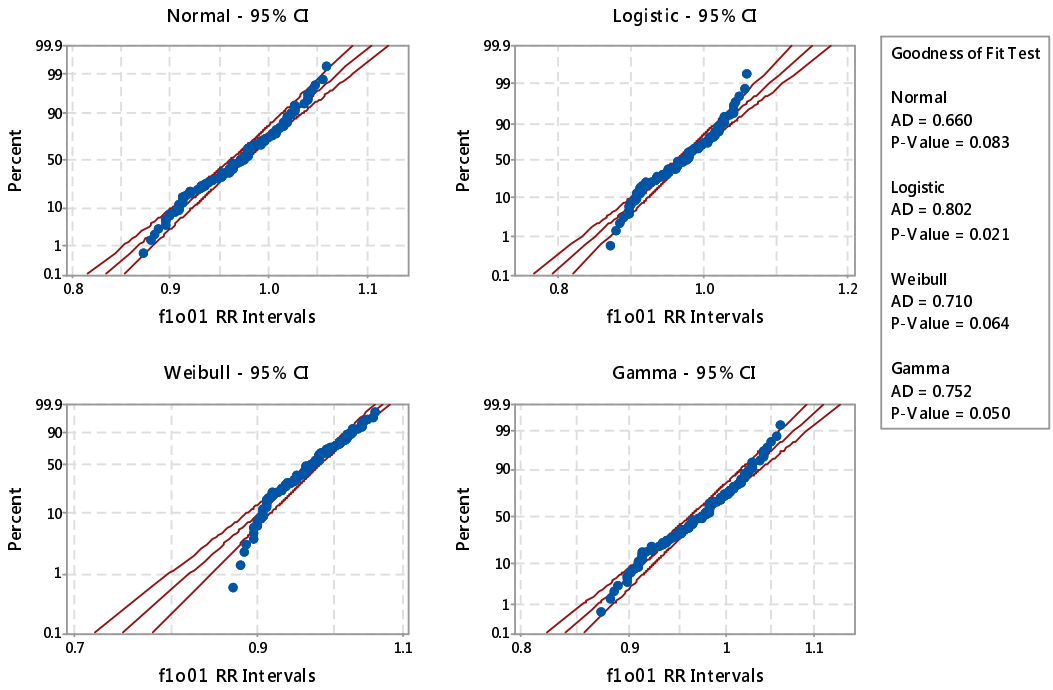

Figure 4.3: Two Minute Sample Distribution Identification of Patient f1o01

of the two minute sample distribution identification of Patient f1o01. The results for normal distribution fit show an AD test statistic of 0.660, and a p-value of 0.083. With significance level of 0.05, there is insufficient evidence to reject H_0. Given that the lowest AD test statistic and highest p-value is observed for normal distribution, it can be stated that this distribution is most plausible fit for f1o01's two minute RR interval sample.

The same technique is applied to identifying the distribution for each of the other nine elderly and ten young patient's chosen for analysis.

4.3.2 Results of Distribution Fitting

Table 4.3 and 4.4 gives summary on the best distribution fit in two minute samples for each of the ten elderly and young patient's from the Fantasia dataset. The parameter estimate, conducted using the MLE method, is also provided in each distribution. Param Est. 1 and Param Est. 2 relate to the following parameters respectively: mean and standard deviation (normal),

location and scale (logistic), shape and scale (gamma), and shape and scale (Weibull).

In the 20 distribution identification tests conducted, the applied technique was able to identify 16 distributions successfully in the patient's RR intervals thus a success rate of 80%. The most commonly identified distribution is Weibull (f1o04, f1yo04, f1yo05, f1yo06, f1y08 and f1yo09), while normal is the second most regularly occurring fit (f1o01, f1o06, f1o10, f1yo02 and f1yo07). Gamma distribution and logistic distribution were identified three times (f1o02, f1o08 and f1o09) and two times respectively (f1o07 and f1y01). An asterisk (*) accompanies results where each distribution tested led to rejection of H^0 but, rather than omit such results, the distribution with the highest p-value, namely Weibull and gamma, in comparison with other distributions in the sample tested, is presented.

The primary novelty of the analysis conducted in this chapter is in demonstrating that numerous statistical distributions may be used to describe RR intervals. Although the results complement the work of Mandreker et al. [52], since Weibull comes first for fittings of RR interval data, it would be naive to describe all RR intervals as a Weibull distribution since results report much better fits for logistic, normal and gamma distribution in some patients. Fittings of the gamma distribution correlate with the results of Hashida et al. [9] while findings of the normal distribution can be backed up by the work of Jennings et al. [8]. It is reiterated that the AD test statistic, in place of the more common KS test as employed by Jennings et al. and Mandrekar et al., has been used in this analysis. The AD test comes second only to the SW test in terms of test power [99] thus it is proposed, and demonstrated in the results, that this test statistic serves to produce more accurate distribution results then past research in this area of work.

Finally, an explanation for why the Weibull distribution occurs various times during the analysis is due to this distribution being very versatile in both its shape and scale, in comparison with the other distributions tested.

Table 4.3: Distribution Fit for Elderly Patients

Dataset	AD Value	P-Value	Best Fit	Param Est.1	Param Est. 2
f1o01	0.66	0.083	Normal	0.9692	0.04389
f1o02	0.74	0.055	Gamma	2296.39891	0.00043
f1o03	0.782	0.042	Weibull*	25.17597	0.98128
f1o04	0.501	0.216	Weibull	34.64655	1.23427
f1o05	1.84	<0.010	Weibull*	52.45041	1.04237
f1o06	0.316	0.536	Normal	1.19325	0.02648
f1o07	0.56	0.102	Logistic	0.9827	0.0167
f1o08	0.635	0.098	Gamma	636.92991	0.00123
f1o09	0.628	0.102	Gamma	296.59721	0.00473
f1o10	0.566	0.14	Normal	0.88009	0.0409

This is more apparent when referring to results in Table 4.3 and 4.4 where the asterisk symbol marks the best fit. Although none of the distributions tested proved a good fit in such tests, the result with the highest p-value was Weibull in three out of four cases (f1o03, f1o05 and f1y03). In one instance, gamma was the closest fit possible (f1y10). Given this observation, the next chapter, modelling of heart data, provides a potential solution to how one may model heart data on samples where there is no good fit under the hypothesised distributions.

4.4 Conclusion

This chapter has focused on distribution identification of heart data, namely RR intervals. Identification of the statistical distribution of RR intervals serves the goal of modelling heart rate, a value that may be derived from RR intervals. Distribution identification was conducted in three steps: hypothesis of heart data distribution based on existing literature, parameter estimates using MLE and performing goodness of fit to validate results via the AD test.

Table 4.4: Distribution Fit for Young Patients

Dataset	AD Value	P-Value	Best Fit	Param Est.1	Param Est. 2
f1y01	0.457	0.215	Logistic	0.76866	0.03145
f1y02	0.352	0.464	Normal	0.90868	0.10745
f1y03	0.761	0.046	Weibull*	12.96976	0.94428
f1y04	0.312	>0.250	Weibull	12.24812	1.3771
f1y05	0.165	>0.250	Weibull	26.66953	0.96275
f1y06	0.485	0.232	Weibull	14.87207	1.02397
f1y07	0.434	0.297	Normal	1.11738	0.13236
f1y08	0.381	>0.250	Weibull	14.00713	0.96614
f1y09	0.625	0.101	Weibull	16.20404	0.8534
f1y10	0.904	0.022	Gamma*	232.26905	0.00335

The chapter has hypothesised that RR intervals may be described by four primary distributions: normal, gamma, logistic and Weibull. The choice of normal, gamma and Weibull distributions is based on findings from the literature review including works by [8], [9, 10] and [52]. In addition, this chapter proposes that there is a high likelihood of RR interval being identified as logistic distribution in cases where RR interval displays wider variation. Parameter estimates was conducted using the MLE method, justification for this choice is due to its wide application in statistics along with unbiased results in $n \geq 50$ sample sizes. The AD test was chosen for distribution identification in place of another common test, KS, because it demonstrates greater test power and less limitation in sample data with outliers.

The results from distribution identification of small samples of RR intervals (up to two minutes) prove the validity of the hypothesis. Also, the proposal that RR interval with wider variation may be better identified as logistic rather than a normal distribution is proven. In conducting an analysis of 20 datasets from the Fantasia database, it has been shown that the technique has a success rate of 80%. The most common distribution fit was found

to be Weibull, and second most common was normal. Third was gamma while logistic is least common. The results in this chapter demonstrate that no one single distribution is capable of describing all heart data. Furthermore, it is shown that the AD test statistic is not perfect since four cases of RR intervals analysed resulted in no good fit for the hypothesised distributions. The next chapter describes the modelling methodology applied for RR intervals based on parameter estimates gathered from this chapter along with comparing it to another common technique: mixture modelling as applied by previous researchers including [102, 58].

Chapter 5

Heart Data Modelling

5.1 Introduction

THE literature review of this thesis shows that the three primary techniques used in heart data modelling include normal mixture distribution, Weibull distribution and neural networks modelling. Although each of the related areas of work provide valid solutions to modelling of heart data, it is the opinion of this research that the prime limitation that these three techniques apply generalisation to heart data modelling. In mixture distribution, it is assumed that heart data may be modelled using various mixtures of normal distribution whilst Mandrekar et al. [52] makes the generalisation that all heart data is Weibull distributed. In the context of neural networks, the techniques employed simply offers a form of curve fitting so that the output data will resemble the input data curve as best as the algorithm is capable of. The hypothesis of this chapter is that an individual distribution better describe heart data in comparison with applying generalisation to all heart data

Findings from the previous chapter have shown that it possible to describe each patient's heart data, i.e. RR intervals, using individual distributions. In two minute samples analysed from twenty patient's datasets, the most

common fit is found to be Weibull distribution. The second most common fit is normal distribution whilst gamma is third and logistic the least common. This chapter aims to compare the viability of using individual distributions in comparison with mixture distribution for modelling of RR intervals. The individual distribution used for modelling of each patient's data, along with parameter values, is defined based on results acquired in the previous chapter. Due to limitation in scope, the results are not compared to neural network algorithms since this field of work encapsulates an entirely different discipline of research. Furthermore, the scope of the comparison relates to modelling RR intervals up to a maximum duration of two minute though discussion on modelling longer periods of time is also touched upon briefly in this chapter.

5.2 Modelling Methodology

For each patient of the Fantasia dataset, RR intervals values are modelled for the equivalent of two minutes, using both individual distribution and mixture modelling techniques. Comparison of the two techniques against the actual real-life data is provided via histogram for visual analysis whilst descriptive statistics are also presented to describe the results in quantitative manner. The next section presents a selection of results obtained from this validation process. Modelling of sample data is conducted using R ver. 3.0.1 [155] whilst figures are rendered using Minitab 17. The methodology applied for modelling of individual distributions of RR interval data and modelling via normal mixture distribution is described in the sections which follow.

5.2.1 Individual Distribution Modelling

Modelling of normal, gamma, logistic and Weibull distribution of RR interval data is conducted using the libraries of *rnorm*,*rgamma*, *rlogis* and *rweibull* respectively. Each function is designed specifically to simulate values in the chosen distribution based on the respective Cumulative Distribution Function (CDF) and Probability Density Function (PDF). The example of modelling Patient f1o01's two minute RR intervals is provided in this section. Patient f1o01 has a sample size $n = 123$ for two minute of RR intervals. From the results of identification distribution this patient's RR intervals are considered normally distributed from the Anderson-Darling (AD) test conducted. Estimation of parameters was achieved via the Maximum Likelihood Estimate (MLE) method and results were previously shown in Table 4.3 and 4.4 for the ten elderly and ten young patient's respectively. An example of the command used in R is shown below:

```
# Modelling of 2 minute RR intervals in Patient f1o01
# using normal distribution
set.seed(100)
f1o01 <- rnorm(123,0.9692,0.04389)
```

All results generated from this chapter use a seed value of 100. The purpose of the *set.seed*() method is to set the randomly generator to a specific known state. This enables the exact same results to be produced on any machine, thus experiments may be replicated by other researchers for validation. Appendix D gives each of the R commands used to model all twenty patient's 2 minute RR intervals.

5.2.2 Mixture Distribution Modelling

For mixture distribution modelling, the R library named *mixtools* is used. Similar to the work of Ketchum et al. [58], two normal distribution com-

ponents are chosen for mixture modelling rather than three as proposed by Costa et al. [102]. This choice is justified due to the fact that the RR intervals analysed consists of only two minutes of data, rather than 24 hour periods thus it is proposed that significant variation in values will not generally take place for normal healthy patients in their heart data values.

To elaborate on the concepts of mixture distribution, as previously discussed in the literature review, one assumes that a distribution of data may be modelled by combining a group of distributions with different weightings. A common technique is to only use normal distribution and perform an estimate of parameters for each curve in observed data, a methodology applied by [102, 58]. In other words, generalisation of the data takes place since the use of normal distribution assumes all heart data of all patients is normally distributed. For the estimation of parameters, the Expectation Maximization (EM) algorithm is used as originally described by Dempster et al. [156]. The EM algorithm is capable of finding the MLE of parameters when datasets are incomplete or missing data values [132]. Although the mathematical proof is out with the scope of this research, the EM algorithm has been widely proven to provide accurate parameter estimates in mixture models. Works including [157, 158] provide an in-depth analysis and discussion on the proof of the algorithm.

The R script named *mixtureModel* is implemented to perform both parameter estimates and model the chosen patient's data. Parameter estimate is conducted using the EM algorithm which comes inbuilt with the *mixtools* library. The *normalmixEM* command automatically computes the mean and standard deviation of each k component of the input data which, in this case, is the RR intervals of each patient within the time frame of two minutes. The parameters estimated are calculated using the EM algorithm, and the results which are shown in the table are provided in Appendix E. An example usage of the *mixtureModel* script, assuming *RRintervalData* is a vector of RR intervals, is invoked as follow:

```
# Example vector of RR Interval data
RRIntervalData = c(1.23, 1.12, 1.31, 1.11...)

#Perform EM estimate and automatically produces model of data
model <- mixtureModel(RRIntervalData)
```

The full script used for mixture modelling is also presented in Appendix D.

5.3 Modelling Results

A selection of visual graphs, i.e. histograms, comparing the real-life RR intervals of the chosen patient and modelled data is given. Figure 5.1 and 5.2 gives two examples of elderly patients whilst Figure 5.3 and 5.4 gives two examples of young patients. The x-axis of the histograms provide the RR intervals of the patient in the unit of seconds. The full descriptive statistics for each patient modelled, compared against the real-data, is provided in Table 5.1 for elderly patients and Table 5.2 for young patients. The next section discusses the results in greater detail. Note that descriptive statistics have been rounded to a maximum of three decimal places for ease of readability.

5.3.1 Results Analysis

From Figures 5.1, 5.2 and 5.3 it could be argued that individual distributions provides a better visual representation the real-life data in comparison with mixture distribution. On the other hand, Figure 5.4 demonstrates that mixture modelling provides a greater resemblance in comparison with individual distribution modelling. Analysis on the underlying statistics,

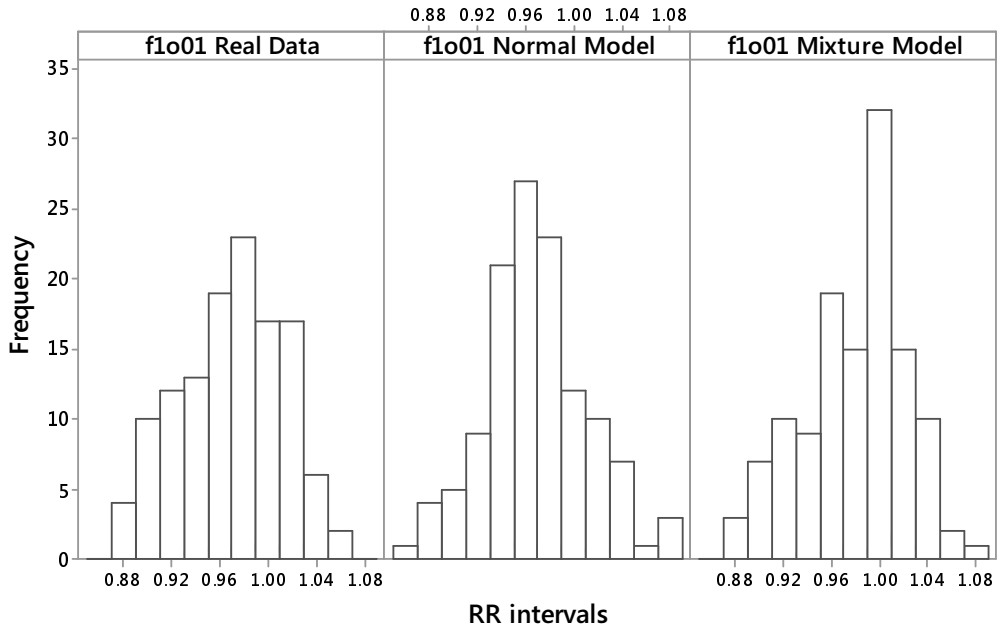

Figure 5.1: Real and Modelled Data Comparison (Patient f1o01)

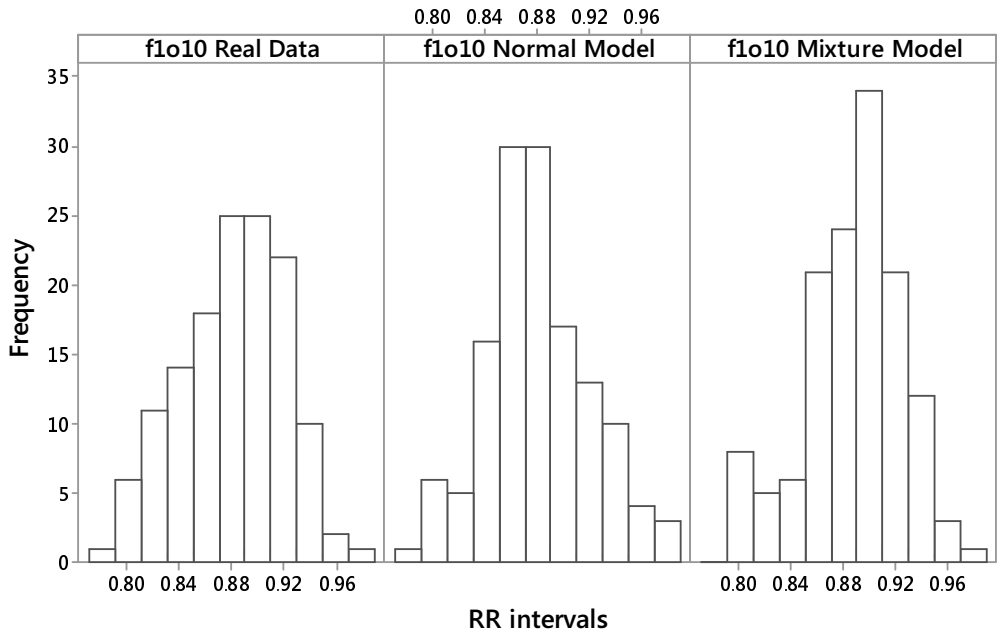

Figure 5.2: Real and Modelled Data Comparison (Patient f1o01)

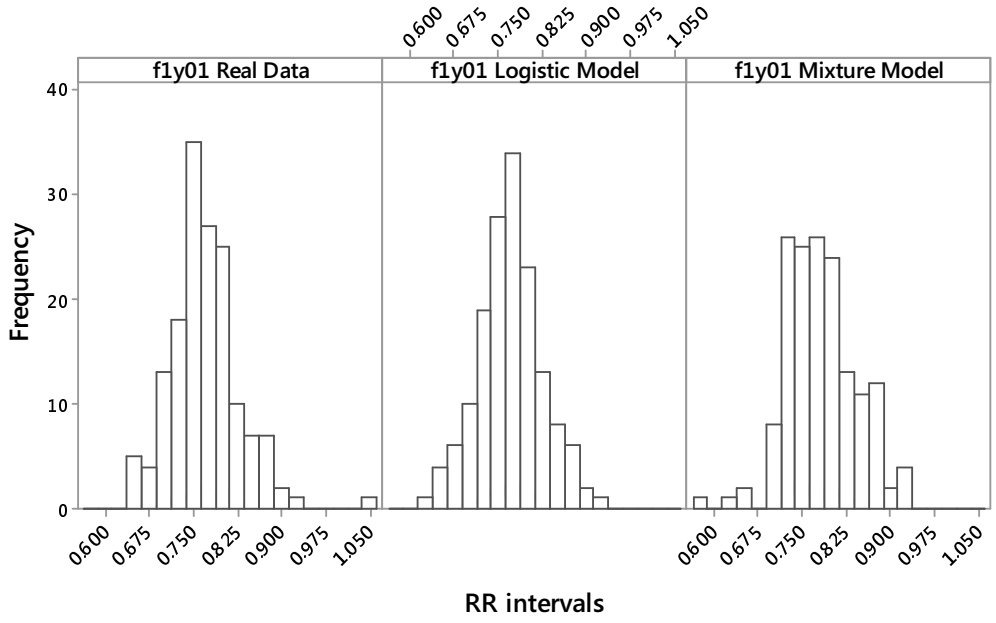

Figure 5.3: Real and Modelled Data Comparison (Patient f1y01)

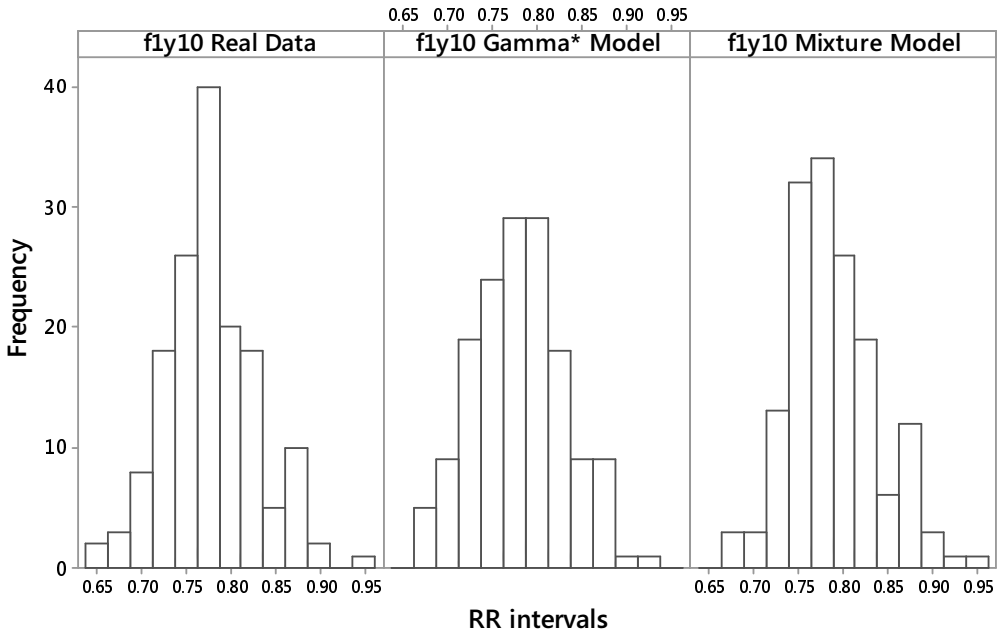

Figure 5.4: Real and Modelled Data Comparison (Patient f1o01)

Table 5.1: Descriptive Statistics on Modelled Data against Real Data (Elderly Patients)

Dataset	Type	Mean μ	SD σ	Median $\mu_{1/2}$	Skewness γ_1
f1o01	Real Data	0.969	0.044	0.972	-0.14
	Normal Model	0.969	0.043	0.966	0.21
	Mixture Model	0.979	0.043	0.987	-0.34
f1o02	Real Model	0.976	0.020	0.976	0.01
	Gamma Model	0.987	0.019	0.989	-0.17
	Mixture Model	0.975	0.019	0.972	0.34
f1o03	Real Data	0.961	0.047	0.972	-0.92
	Weibull* Model	0.959	0.043	0.967	-0.82
	Mixture Model	0.959	0.039	0.964	-1.12
f1o04	Real Data	1.215	0.041	1.218	-0.33
	Weibull Model	1.212	0.038	1.220	-1.05
	Mixture Model	1.212	0.043	1.212	-0.18
f1o05	Real Data	1.031	0.026	1.040	-0.81
	Weibull* Model	1.030	0.023	1.034	-0.93
	Mixture Model	1.034	0.024	1.039	-1.09
f1o06	Real Data	1.193	0.027	1.196	-0.24
	Normal Model	1.194	0.027	1.192	0.15
	Mixture Model	1.196	0.025	1.203	-0.57
f1o07	Real Data	0.981	0.031	0.984	-0.56
	Logistic Model	0.985	0.027	0.984	0.09
	Mixture Model	0.985	0.031	0.992	-0.91
f1o08	Real Data	0.781	0.031	0.776	0.29
	Gamma Model	0.785	0.030	0.786	0.09
	Mixture Model	0.786	0.031	0.781	0.38
f1o09	Real Data	1.402	0.082	1.388	-0.03
	Gamma Model	1.402	0.079	1.402	-0.03
	Mixture Model	1.407	0.078	1.423	-0.43
f1o10	Real Data	0.880	0.041	0.884	-0.15
	Normal Model	0.880	0.040	0.876	0.21
	Mixture Model	0.887	0.039	0.894	-0.42

Table 5.2: Descriptive Statistics on Modelled Data against Real Data (Young Patients)

Dataset	Type	Mean μ	SD σ	Median $\mu_{1/2}$	Skewness γ_1
f1y01	Real Data	0.771	0.058	0.764	0.84
	Logistic Model	0.769	0.055	0.768	0.1
	Mixture Model	0.781	0.060	0.775	0.06
f1y02	Real Data	0.909	0.107	0.906	0.03
	Normal Model	0.907	0.105	0.900	0.2
	Mixture Model	0.911	0.098	0.898	0.17
f1y03	Real Data	0.907	0.085	0.920	-0.44
	Weibull* Model	0.903	0.082	0.920	-0.83
	Mixture Model	0.922	0.080	0.935	-0.47
f1y04	Real Data	1.321	0.126	1.338	-0.3
	Weibull Model	1.312	0.114	1.325	-0.9
	Mixture Model	1.336	0.115	1.338	-0.04
f1y05	Real Data	0.943	0.044	0.948	-0.87
	Weibull Model	0.942	0.040	0.951	-0.82
	Mixture Model	0.945	0.036	0.948	-0.49
f1y06	Real Model	0.989	0.080	0.992	-0.48
	Weibull Model	0.980	0.087	0.985	-0.6
	Mixture Model	0.998	0.080	1.014	-0.55
f1y07	Real Data	1.117	0.132	1.120	-0.07
	Normal Model	1.116	0.134	1.108	0.19
	Mixture Model	1.157	0.126	1.186	-0.41
f1y08	Real Data	0.931	0.081	0.948	-0.57
	Weibull Model	0.926	0.079	0.940	-0.83
	Mixture Model	0.933	0.077	0.954	-0.59
f1y09	Real Data	0.827	0.060	0.836	-0.52
	Weibull Model	0.824	0.060	0.835	-0.82
	Mixture Model	0.826	0.049	0.836	-0.48
f1y10	Real Data	0.778	0.052	0.772	0.41
	Gamma* Model	0.780	0.050	0.781	0.12
	Mixture Model	0.787	0.050	0.777	0.54

provided in Table 5.1 and 5.2 shows deeper insight on comparison between individual modelling and mixture modelling.

The descriptive statistics presented include mean, median, standard deviation and skewness of the modelled data. Overall, it is observed that the mean of individual distribution models better match the real-life RR interval values. In total, thirteen individual distribution models have a closer mean value to the real-life data in comparison mixture modelling whilst four modelled patient's resulted in equal values. Thus it can be stated overall that the use of a individual distribution is better for capturing the average values of RR intervals in comparison with mixture modelling.

Similarly, thirteen cases of individual distribution provide closer values in median to the real-life data in comparison with mixture modelling. For standard deviation, ten of the modelled data using individual distribution provide a more accurate representation of this measurement whilst five of the patient's modelled provided equal results with mixture modelling. On the other hand, mixture modelling appears more accurate for skewness where thirteen results produced more accurate results in comparison with individual distribution.

In the case of real-life patient data identified as normal distribution, the descriptive statistic shows that the use of the individual distribution technique, i.e. modelling of a single normal distribution based on estimated parameters, provides the most accurate results in regards to mean, median and standard deviation. In the case of logistic modelling, it is found that both individual and mixture modelling are equally valid. However, this finding is limited due to the fact that only two patient's RR intervals within the datasets analysed show evidence of this distribution.

For patient's heart data identified as gamma distribution, the use of this individual distribution proved less effective than mixture modelling in two primary cases: Patient f1o02 and f1o08. Although the mean, median and

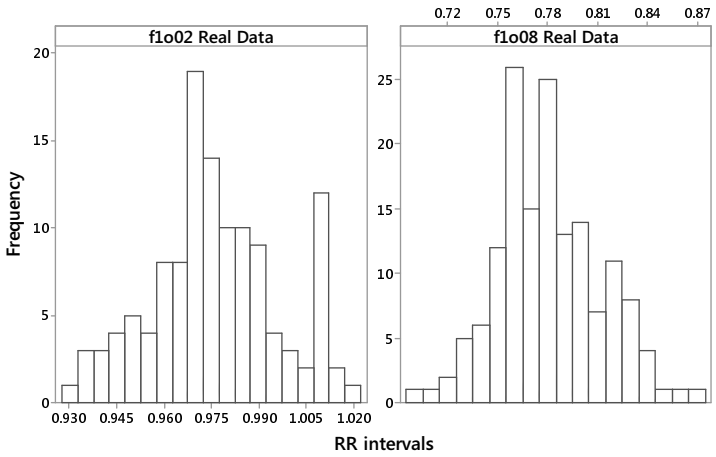

Figure 5.5: Two Peak RR Interval Distribution (Patient f1o02 and f1o08)

standard deviation are quite similar, the shape of these two models are more accurately represented by mixture modelling. Analysis on the real-life dataset provides an answer as to why mixture modelling works better. In both cases, it is found that two distinct peaks are found in both patient's RR intervals as shown in Figure 5.5. Thus, this confirms bimodal characteristics in the RR interval data confirming the results of [11] and, although statistical evidence suggests that these two datasets are gamma distributed, mixture distribution may be better suited to modelling the characteristic of multiple peaks in heart data. In regards to the Weibull distribution, it is found that the use of this individual distribution for modelling is equally comparable to mixture modelling. Both techniques provide similar values in regards to the real-life patient data but it is found that mixture modelling better captures the skewness of the data once again.

Results for Weibull* models - marked with the asterisk symbol - which, as discussed previously in Section 4.3.2, relate to fits where a p-value < 0.05 therefore meaning no good fit exists for the data (though the presented distribution is deemed to be the closest fit in comparing p-values) are also similar with mixture modelling. In two cases (f1o03 and f1yo03), the use of Weibull* results in a better representation of the real-life patient data whilst

the third case (f1o05) results in more accurate mean and skewness in comparison with mixture modelling. Despite the lack of significant evidence that this distribution is the correct fit for these three patients, the findings here demonstrate the flexibility in the Weibull distribution in modelling of RR interval data and validity of this distribution for modelling of heart data as discussed in the work of [52]. Finally, in regards to the single Gamma* model, the finding shows that mixture modelling is the ideal technique rather than using an individual distribution which better correlates with the analysis conducted in the previous chapter since no good fit was found for this sample.

It can be summarised that individual distribution modelling is preferable in the case of evidence that RR intervals can be described using a single distribution. On the other hand, mixture modelling is preferable to modelling RR intervals where there is insignificant evidence to demonstrate that the real-life data samples analysed follow a specific distribution. Furthermore, mixture modelling is ideal in the modelling of bimodal characteristics found in RR interval data. Finally, given the fact that no good fit is produced for RR interval's of greater than two minutes (see Section 4.3), in circumstances where longer periods of RR interval need to be modelled, mixture modelling proves to be a more useful tool than individual distribution modelling as demonstrated in previous works including [102, 58].

5.3.2 Summary and Modelling Recommendations

From the results, the following points may be summarised when comparing individual distribution against mixture modelling of RR interval data:

- Use of individual distribution better represents the mean, median and standard deviation of a patient's RR interval data in comparison with mixture modelling.

- Use of mixture distribution better encapsulates the skewness, i.e. asymmetry, of RR intervals due to the use of two normal distributions. Since two means and two standard deviations are defined, a positive or negative skew can be easily produced in modelling patient's RR intervals which demonstrate skewness.
- Individual distribution modelling is ideal when there evidence that the real-data values follow a specific distribution (as identified using the AD test in the previous chapter).
- Mixture modelling proves to be valuable in modelling RR intervals when the patient's data does not fit a specific distribution or demonstrates bimodal characteristics.
- Mixture modelling is ideal for modelling longer periods of RR interval whilst individual distribution is ideal for short periods of RR interval which fit specific distributions.

Based on these findings, Figure 5.6 presents a simple decision flow chart of the modelling process for RR interval data which may be of use for future research into this area.

5.4 Discussion

Modelling of RR interval data for this chapter has been conducted and presented in the frequency domain. The modelled data has been presented in a series of histograms and descriptive statistics thus showing both quantitative and qualitative validation of both individual distribution and mixture modelling. However, one primary limitation still exists in the results presented in this chapter. This limitation relates to the lack of realistic behaviour on RR intervals modelled. This becomes more apparent in graphing a time-series plot of the first minute sample of Patient f1o01's RR interval data against the modelled data as shown in Figure 5.7. It should

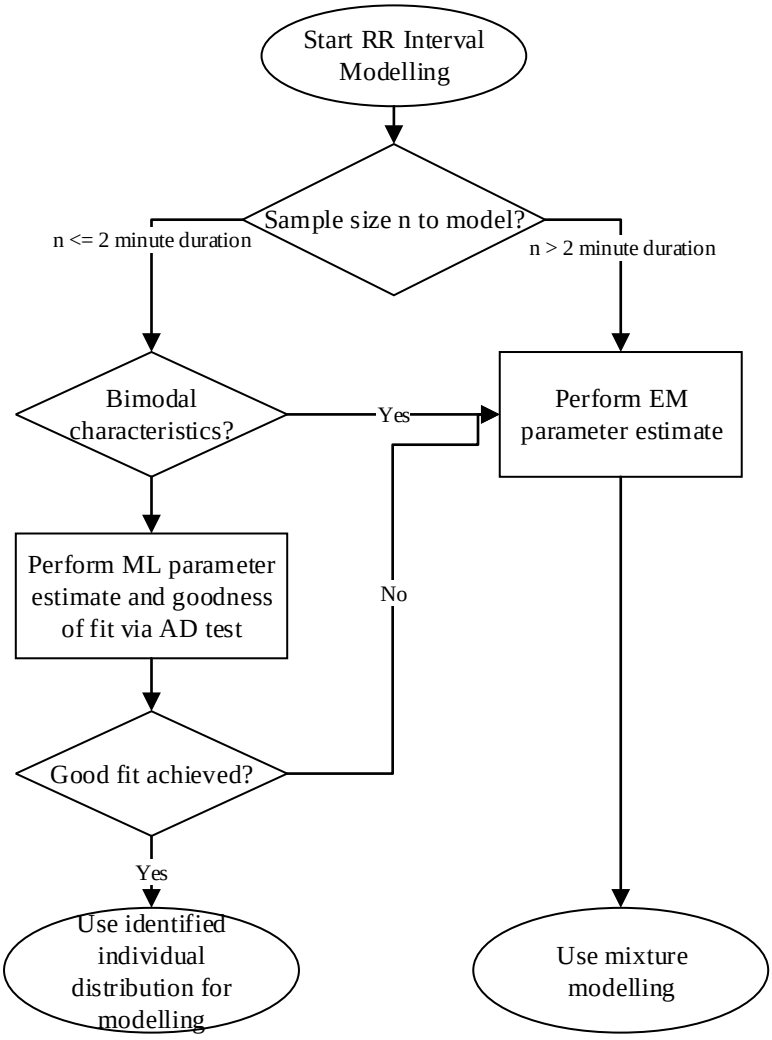

Figure 5.6: RR Interval Modelling Process

be noted similar results are observed in using the technique of mixture modelling. In the time-series plot, real life data of the patient's RR interval exhibits a more controlled fluctuation of peaks and nadirs whilst both modelled data has a distinctly sharper fluctuation at any given time interval. In other words, the modelled data does not have any control over the behaviour of variation as seen in the real-life data.

Research was conducted in this area of work to determine whether it is possible to accurately model heart rate variation (HRV) in the time domain but results proved inconclusive. From the research carried out, there is a lack of recent work which attempts to accurately model and simulate the behaviour of HRV within the time domain. The work which does exist, including McLernon et al. [159] and Georgieva-Tsaneva et al. [160], claim to present mathematical models which are capable of providing control, and realistic simulation, on heart flunctuations. However, neither work provides any form of validation (as Law [42] strongly emphasises for achieving successful simulation) therefore evidence of realism, in comparison with real-life data, is missing.

It is proposed that one primary reason for the overall lack of research in this area may relate to the non-linear dynamics of HRV which leads to the concept of whether the heart may be considered a chaotic system. Prior to the conclusion to this chapter, this section provide an overview on how the concept of chaos theory is related to HRV to provide justification for the limitation found in the modelling of RR intervals. This section may also prove to provide areas of future work for other researchers in the field of heart data modelling.

5.4.1 The Chaotic Heart

The concept of chaos theory was first discussed by Jules Poincaré and further formalised by Edward Lorentz. In short, chaos theory states that a

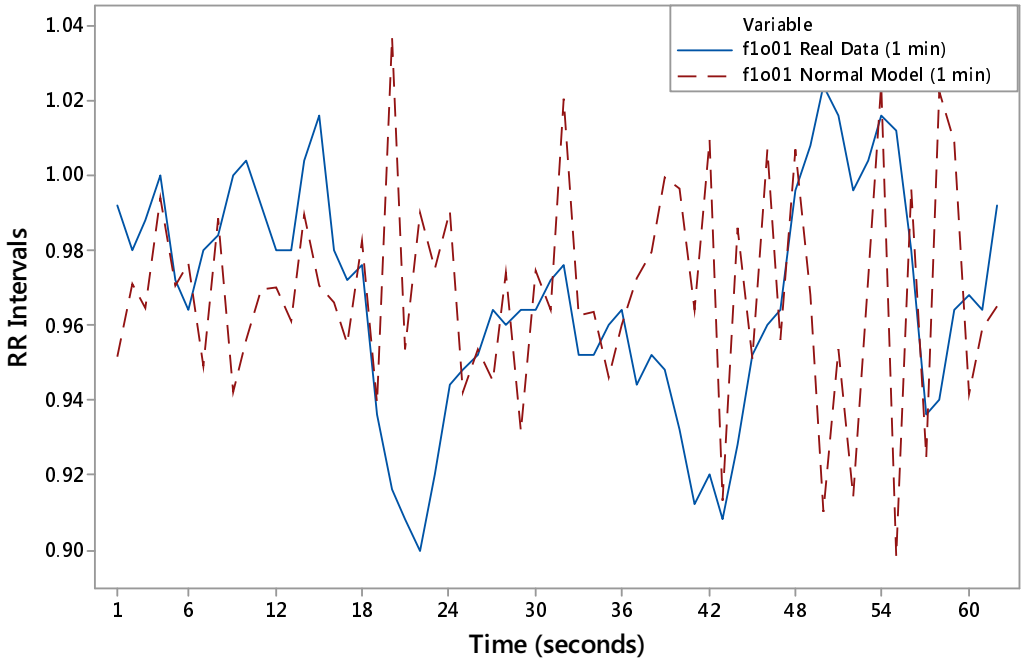

Figure 5.7: Time-Series Comparison (Patient f1o01)

dynamic system is considered chaotic if it is widely sensitive to initial conditions. Chaotic systems, in theory, are considered deterministic but due to its characteristic of sensitivity to initial conditions will produce results which are widely unpredictable over longer periods of time. An example of this is weather forecasting systems which have capability of only predicting the weather for a certain period of days before the results are widely inaccurate. This is due to initial conditions, e.g. temperature, atmospheric pressure, wind speed etc., which can only be submitted to the system with a certain degree of accuracy.

Numerous discussions and publications have been produced in determining whether the heart itself is a chaotic system. Two of the main concepts in this area of research relates to the *fractal* shape of heart rate variability and the characteristic of *chaos* in the behaviour of heart rate within the time domain [161]. Fractal relates to the general shape of heart variability, which should have a self-similar pattern when analysed at different scales, whilst the term chaos relates the actual behaviour of the heart's fluctua-

tion whereby it follows a nonlinear but, in theory, deterministic pattern. It has been proposed that analysis of the chaotic nature of the heart not only provides greater insight into the inner workings of this organ but may also help to better identify diseases - especially in regards to age related conditions [162, 161]. Similar view is shared by Lombardi [153] and, more recently, Krstacic et al. [163] in which both authors observe non-linear characteristics in HRV and propose that abnormalities [153] and heart failures [163] may be detected depending on the fractal dynamics of HRV.

On the opposite side of the spectrum, research has also been published which questions the validity of whether the heart is considered chaotic. In the analysis of 10 healthy patients, Kanters concludes that there is no evidence of chaos in RR interval data [164]. More recently, as stated by Glass [165], the author proposes the question to ask is not whether the heart displays chaotic behaviour but rather *"What are the mechanisms underlying complex cardiac rhythms and how are they manifest in the laboratory and clinic?"* [165]. The primary point of this statement relates to the fact that Glass believes the underlying dynamics of the heart are not chaotic but simply have yet to be explained within research. In other words, not enough data exists to give us a full picture of exactly how the heart functions.

Despite evidence for and against chaos in the heart, it can be made clear that many factors will affect fluctuation in HRV. As discussed previously in this thesis, simple factors include the level of activity in a person, general fitness level [153], posture [154] and external stimuli [10]. If these factors are considered as initial inputs to a heart rate modelling system, one can already see the difficulty in the system producing an accurate output at any given moment of time due to the need to correlate such variables along with deducing such variables into values which an algorithm is capable of interpreting, e.g. how does one define the posture of a person using an accurate value? Thus, it's proposed that attempting to accurately model the underlying behaviour of HRV is currently out with the scope and possibil-

ity of this research. Further work must be conducted in proving if the heart itself is chaotic, and if so, a technique must be designed in order to accurately model heart rate data's pattern over periods time whilst considering the vast multitude of variables which may potentially affect the fluctuations of the heart. Complimentary works which are relevant to the concept of chaos in the heart, along with biology in general, include [166, 167, 168].

5.5 Conclusion

This chapter hypothesised that an individual distribution better describe heart data in comparison with applying generalisation technique. The successful results of modelling RR intervals using an individual distribution, based on parameters calculated from MLE and evidence of underlying distribution via the AD test statistic, proves that a simple technique for producing RR intervals which are similar to real-life data is possible. The results in modelling twenty patient's from the Fantasia dataset show that individual distribution is suited to modelling RR intervals when there is strong proof to show the real-life samples follow a specific distribution.

On the other hand, comparison of individual distribution modelling against normal mixture distribution showed that when real-life RR intervals cannot be correctly identified, i.e. no good fit with any distribution, mixture modelling proves to a be a better choice of technique. Additionally, it is acknowledged that mixture modelling is a better technique for modelling the bimodal characteristics found in some patient's data. Furthermore, as modelling was been conducted on small samples of patient's RR interval, it's proposed that mixture modelling provides a better approach when sample size is greater than two minutes. Based on these findings, a simple decision flowchart was provided to assist in decision making when attempting to model RR intervals for future research (Figure 5.6).

The main limitation in this chapter is modelling of variation between each

RR interval. A discussion has been provided which shows that there is potential in the heart being a chaotic system thus providing justification in the difficulties in accurately modelling such behaviour due to the vast variety of inputs required and the heart's potential sensitivity to initial conditions. Thus, although modelling of the heart data within the time domain is still an unsolved area of research, frequency distribution analysis and modelling has many uses including further understanding on how the heart works along with applications within healthcare technologies. In this research, focus is given on the latter, and the next chapter demonstrates the applicability of using simulated heart rate, derived from the modelled RR intervals using both individual and mixture distribution, for the context of demonstrating evidence towards improved accuracy in the novel Early Warning Score (EWS) risk assessment algorithm proposed.

Chapter 6

Heart Data Risk Assessment

6.1 Introduction

THE previous chapters have shown that it is possible to identify single distributions in small samples of heart data through statistical analysis and such data can then be modelled via an individual probability distribution. In heart data samples, i.e. RR intervals, whereby no good fit exists, it has been shown that mixture distribution is capable of modelling such data. Having demonstrated the techniques in heart data modelling, this chapter brings forwards these concepts and shows how such data may be applied in scenarios on the Early Warning Score (EWS) algorithm. In particular, by modelling RR intervals and deriving instantaneous heart rate (IHR) for use in simulation scenarios, this chapter demonstrates a potential improvement to the EWS algorithm using quartile-based statistics - an anomaly based Intrusion Detection System (IDS) technique which enables customisation of risk thresholds for each individual patient analysed. The quartile-based implementation aims to address one of the noted limitations of the standard EWS algorithm, the issue of sensitivity as discussed by Gao et al. [129]. Evidence of this sensitivity is given in this chapter by evaluating the standard EWS algorithm against normal healthy patient's heart data. Although it is acknowledged that the EWS algorithm

considers multiple vital signs, for the scope of this chapter's experiments focus specifically on risk assessment of heart data only.

6.2 Experiments Overview

Two minute samples of modelled heart data are used to evaluate the sensitivity in different approaches for risk assessment (Figure 6.2). Simulation of RR intervals is achieved by first modelling the values based on the Fantasia dataset using a probabilistic approach. The act of simulation itself is performed in a discrete manner. Presence of time in the simulation, as described by [31], is not considered in the experiments conducted in this chapter since the risk assessment algorithm's evaluated does not consider this parameter during analysis. Modelled heart data values are generated using the seed values provided in Appendix D.1 for ease of reproducibility.

Individual distribution and normal mixture distribution are the two modelling techniques used for modelling of heart data. From results obtained from Chapter 4, the individual distribution method is used for modelling of patient's heart data when identification of distribution was successful. For patient's heart data that did not provide sufficient evidence to fit a particular distribution, the normal mixture modelling method is used. Specifically, Patient's f1o03, f1o05, f1y03 and f1y10 are modelled using normal mixture modelling while all others are modelled based on their individual identified distribution. Although it is acknowledged that mixture modelling provided better replication of bimodal characteristics in some of the patients, overall it is found that individual distribution more accurately modelled mean, median and standard deviation. Mean and median, in particular, are values that are deemed more necessary for evaluating the accuracy of the EWS algorithm since the logic of this system calculates risk on a value-by-value basis rather than considering the shape of the data.

The first experiment applies the modelled data against an electronic ver-

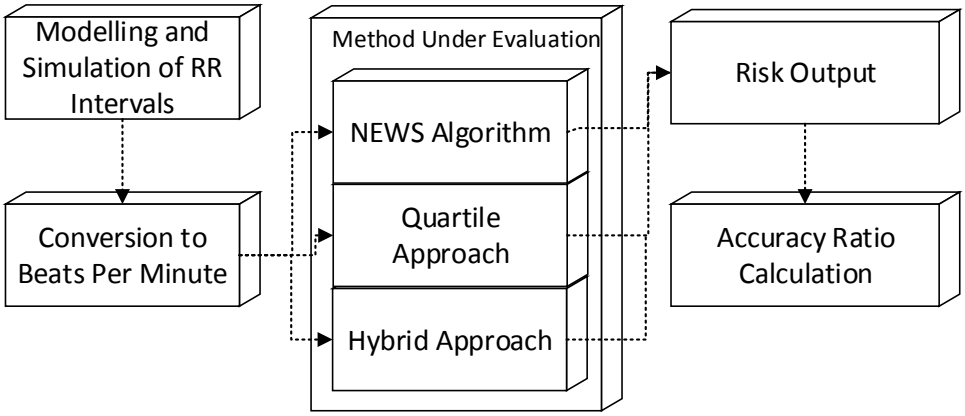

Figure 6.1: NEWS, Quartile and Hybrid Experiment Work Flow

sion of the National Early Warning Score (NEWS) algorithm while the second experiment demonstrates the capabilities of an anomaly-based approach via quartile statistics. The third experiment presents a hybrid approach, which integrates the knowledge-based method, i.e. the NEWS algorithm, with the anomaly-based approach, i.e. quartiles statistics, to demonstrate the improved accuracy of this technique.

6.2.1 Metrics of Evaluation

Two primary sets of metrics are presented in this experiment: heart rate risk scores (1, 2 and 3) and accuracy ratio. Any risk score of 1 or greater is considered a false-positive in this experiment. This is due to the fact that the dataset used consists of recordings of normal health patients. The accuracy ratio is a metric presented in this thesis which represents the overall sensitivity of the EWS algorithm when assessing normal health patient's heart rates.

The accuracy ratio metric is inspired and based upon work conducted by Sommers et al. [13] and Rab and Kalam [14] on evaluation of IDSs. As demonstrated in the literature review, Sommers et al. proposes the metric of $Efficiency = \frac{TruePositives}{AllAlarms}$ [13] while Rab and Kalam provide the metric of

$FAR = \frac{FalsePositives}{AllAlarms}$ [14]. The review has shown that FAR is simply derived from $1 - Efficiency$, in other words it can be considered the opposite of the $Efficiency$ metric. In metrics defined for evaluation of IDS, the prior assumption is made that testing will involve both malicious and normal activity, thus true-positives and false-positives have the potential to occur respectively. In other words, the variable $AllAlarms$ derives its meaning from inclusion of all forms of alarms, i.e. true-positives + false-positives + false-negative, but in the case of testing of the EWS algorithm, using normal healthy heart datasets, there is only potential for false-positives to occur thus resulting in traditional IDS metrics providing little information except from the fact that a value of 0 means all triggers are false-positives [13].

Given the issue faced with traditional IDS metric for measurement of false-positives, modifications are made to the existing equations and the accuracy ratio in this thesis is presented as:

$$AccuracyRatio = 1 - \frac{FalsePositives}{\sum_{i=1}^{n} s_i} \tag{6.1}$$

$FalsePositives$ is the number of total false-positives produced by the EWS algorithm during experimentation while s_i is simply the total number of RR intervals data samples under test. As with the work of Sommers et al., a value of 1 in this metric shows that no false-positives occur in the EWS algorithm under test, therefore 100% accuracy, whilst a result of 0 shows that all values in the dataset resulted in false-positives (0% accuracy).

6.3 NEWS Approach

This section aims to demonstrate the argument put forth by [129, 131] in that the EWS algorithm is limited due to sensitivity, and does not provide enough evidence to demonstrate positive or negative outcome in a patient's health. The proposed standard for NHS hospitals, namely the

NEWS algorithm, is evaluated against modelled heart data. The background chapter has provided the full set of parameters monitored in the NEWS algorithm, thus Table 6.1 only gives the heart rate values, and associated risk score for ease of reference. Although the literature review has shown that the NEWS and MEWS algorithm are the two most common Track and Trigger (TT) systems employed in clinical environments, preliminary experiments showed that little variation existed in results for these two algorithms thus focus is given on the NEWS algorithm. Appendix F provides further evidence on the minor variation of results between the NEWS and MEWS algorithm when evaluated against modelled heart data. An electronic implementation of the NEWS algorithm for assessment of heart data was achieved in R ver. 3.0.1, and the methodology is discussed in the next section.

6.3.1 Methodology

As noted earlier, the electronic version of the EWS system, regardless of algorithm, is simply a set of logical conditions making it simple to migrate to an electronic system. An example of the logic to determine if heart rate is within 41 to 50 Beats Per Minute (BPM), written in pseudocode, is given below:

```
ALERT
IF HeartRate <= 50 AND HeartRate >= 41
```

If this example statement is true, then the alert produced would be a risk score of 1. Each risk score result is automatically logged in the electronic

Table 6.1: NEWS Algorithm for Heart Rate Risk Scores

Parameter	**3**	**2**	**1**	**0**	**1**	**2**	**3**
NEWS Heart Rate	$\leq$40		41-50	51-90	91-110	111-130	$\geq$131

NEWS algorithm, and the metric of accuracy derived from the total results.

Each modelled RR interval, derived from the Fantasia dataset, is fed into the electronic EWS, in a discrete manner, and the NEWS algorithm's risk score for heart rate is automatically computed. All twenty modelled patient's two minute samples from the Fantasia dataset are used for this experiment. As previously explained, the Fantasia dataset consists of ECG recordings of elderly and young patient's, in a resting state. The primary aim of the original study, conducted by [147], was to assess variations in heart rate between healthy young and elderly individuals. The datasets were manually reviewed by experts in this field of research to ensure the recordings did not consist of any abnormalities. Thus, it can be justified that models of RR interval derived from this dataset consists of heart data which is within the scope of normality. Since the EWS systems require heart data in the unit of BPM, the RR intervals are automatically converted to this unit using the equation $IHR = \frac{60}{RR^{interval}}$ as originally presented in the Theory chapter.

The full implementation of the electronic NEWS algorithm for risk assessment in heart data is implemented using R and full source code is given in Appendix D.5.

6.4 Quartile-Based Approach

The review on statistical-based anomaly detection techniques has shown that the prime advantage of this approach is that there is no prior assumpting on the distribution of the data in which it is applied to. This compliments risk assessment in heart rate since the analysis and modelling from earlier chapters has shown that although there is a 80% success rate on distribution identification of RR intervals - values in which IHR may be derived from. In other words, there is a wide variety of distributions which heart data follows thus making prior assumptions on probability distri-

bution a research intensive task. The implementation of a quartile-based risk assessment system for heart rate is primarily inspired by the anomaly detection method described by [16] in which the technique was applied successfully to detect anomalies in data centres.

6.4.1 Methodology

The methodology for the quartile-based approach to risk assessment in heart data can be considered to encapsulate two primary steps: 1) training phase and 2) evaluation phase (Figure 6.2). The training phase can be considered similar to the techniques applied during the DARPA evaluation [12] whereby benign datasets are analysed by the anomaly detection system under evaluation to learn the concept of normal behaviour. Training, in this experiment, is achieved using a quartiles learning approach implemented in R whilst the datasets used included twenty normal healthy patient's RR interval found in the Fantasia database.

Once the training phase is completed, simulation of the twenty patient's two minute samples takes place in the same manner as applied to the NEWS approach. For the assessment of risk in the simulated patient's heart data, a rule-based technique is defined for the quartile approach. The quartile approach consists of three primary rule set which triggers an alert if certain upper or lower thresholds for a specific patient's heart data value is produced. Thus, an example rule using pseudocode is defined as:

`ALERT`

`IF HeartRate >` $UT_{LowRisk}$ `AND IF HeartRate <=` $UT_{MedRisk}$

In other words, a low risk is defined as any value of heart rate which is greater than $UT_{LowRisk}$ but less than or equal to $UT_{MedRisk}$. Furthermore, similar to the NEWS approach, risk scores are automatically computed to allow for calculation of the accuracy ratio metric. The steps taken for cal-

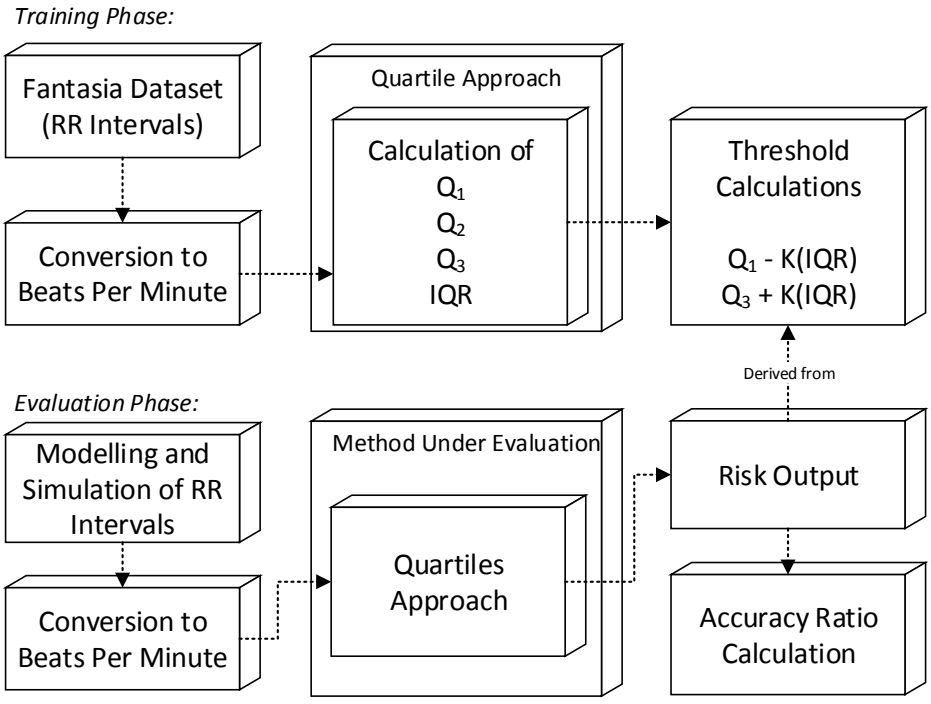

Figure 6.2: Quartiles Based Experiment Methodology

culation of quartiles in the training datasets along with definition of risk classification based on threshold values calculated is presented in the next section.

6.4.2 Calculation of Anomalies

Quartiles is the statistical technique of defining data range as 25% components and has been traditionally used for generation of box plots as show in the work of [15]. The primary attributes of quartiles are Q_1, Q_2, Q_3 and the IQR (Interquartile Range). The 25th percentile of data is represented by Q_1, the 50th percentile by Q_2, i.e. the median, and the 75th percentile by Q_3. The IQR is the difference between Q_3 and Q_1. Calculations of quartiles, using the method as define by Tukey [15] is conducted in the following steps

1. Sort all data into ascending numerical order.

2. Calculate Q_2 which is the median of the whole sample. If data size n is even, refer to Equation 6.2 else refer to Equation 6.3.

3. Calculate Q_1 which is the median of the lower half of Q_2.

4. Calculate Q_3 which is the median of the upper half of Q_2.

5. Calculate IQR, the difference between Q_3 and Q_1, which is represented in Equation 6.4.

$$Q_2 = \frac{\frac{n}{2}th \text{ item} + \frac{n}{2+1}th \text{ item}}{2}th \text{ item} \tag{6.2}$$

$$Q_2 = \frac{n+1}{2}th\ item \tag{6.3}$$

$$IQR = Q_3 - Q_1 \tag{6.4}$$

Having calculated the quartiles, anomalous data can then be defined based on the value derived for IQR. In the standard quartiles approach, as described by Tukey, classification of anomalous data comes in two forms of thresholds: *"outside"* and *"far out"* [15]. The work of Wang et al. [16] refer to these two thresholds as *"possible anomalies"* and *"anomalies"* [16]. In both works, the lower and upper thresholds for anomalous data is calculated using $Q_1 - K(IQR)$ and $Q_3 + K(IQR)$ where K is the values of 1.5 and 3.0 respectively. In other words, a possible anomaly is any variable greater than the upper or lower half of median multiplied by one and half times the IQR whilst anomalous data is considered any variable greater than three multiplications from the same calculation.

This thesis uses the terminology as applied by the EWS system for classification of anomalous data which is referred to as *low risk* and *high risk* for *"possible anomalies"* and *"anomalies"* respectively. Furthermore, a third threshold is introduced in this thesis, referred to as *medium risk* which sits

between the prior two thresholds. Thus $K = 2.25$ for *medium risk*. Calculation of low risk, for lower and upper threshold of data (defined as LT and UT), is achieved as follows:

$$LT_{LowRisk} = Q_1 - 1.5(IQR) \tag{6.5}$$

$$UT_{LowRisk} = Q_3 + 1.5(IQR) \tag{6.6}$$

whilst calculation of medium risk is achieved as follows:

$$LT_{MedRisk} = Q_1 - 2.25(IQR) \tag{6.7}$$

$$UT_{MedRisk} = Q_3 + 2.25(IQR) \tag{6.8}$$

finally, calculation of high risk is calculated as:

$$LT_{HighRisk} = Q_1 - 3(IQR) \tag{6.9}$$

$$UT_{HighRisk} = Q_3 + 3(IQR) \tag{6.10}$$

In other words, a possible low risk in a patient's heart data is any value which is 1.5 times away from the IQR whilst a high risk is considered any value which is 3 times away from the IQR. A medium risk is any value which is 2.25 times away from the IQR. Classification of normality is defined as any variable within the limits of $Q_1 - 1.5(IQR)$ and $Q_3 + 1.5(IQR)$. Table 6.2 presents a comparison of these three classifications of risk in comparison with the NEWS algorithm for heart data. As it can be seen, the primary difference between the quartile-based approach and the NEWS algorithm is the fact that the quartile-based will derive the risk of a patient's

Table 6.2: Comparison of Risk Score in NEWS and Quartile Based Approach

Parameter	**3**	**2**	**1**	**0**	**1**	**2**	**3**
NEWS Heart Rate	≤ 40		41-50	51-90	91-110	111-130	≥ 131
Quartile Approach	$< Q_1 - 3(IQR)$	$\geq Q_1 - 3(IQR)$ and $< Q_1 - 2.25(IQR)$	$\geq Q_1 - 2.25(IQR)$ and $< Q_1 - 1.5(IQR)$	$\geq Q_1 - 1.5(IQR)$ and $\leq Q_3 + 1.5(IQR)$	$\leq Q_3 + 2.25(IQR)$ and $> Q_3 + 1.5(IQR)$	$\leq Q_3 + 3(IQR)$ and $> Q_3 + 2.25(IQR)$	$> Q_3 + 3(IQR)$

heart data based on learning of past values whilst the NEWS algorithm is completely static.

An implementation of the quartiles approach is achieved using R. The script accepts, as input, a data vector and automatically calculates the Q_1, Q_2, Q_3, IQR. Customised $LT_{LowRisk}$, $UT_{LowRisk}$, $LT_{MedRisk}$, $UT_{MedRisk}$, $LT_{HighRisk}$ and $UT_{HighRisk}$ for each individual patient is then computed. Appendix D.3 provides source code on the training script whilst Appendix D.4 gives the source code to decision making logic.

6.5 Hybrid Approach

The literature review has shown that, in the context of Clinical Decision Support System (CDSS), two primary categories of approaches are applicable: knowledge-based techniques and machine-learning techniques. Knowledge-based techniques are widely more applicable in live healthcare environments due to lesser complexity in understanding of the system along with the fact that knowledge-based approach provides greater transparency in the decision making process [110]. It's proposed that the quartiles based approach, which may fall under the context of machine-learning, does not share the limitation of complexity, nor transparency due to the relatively simple mathematics involved. However, it is limited in the fact that no expert clinical input has been provided in calculation of normal data and potential outliers. Thus, this section proposes an integration quartiles based approach with the NEWS algorithm for risk assessment in heart data, with the goal of improving accuracy whilst retaining transparency in the clinical decision making process.

6.5.1 Methodology

Similar to quartile approach, each patient's real-life heart data of the Fantasia dataset is placed through a training process. The attributes of *LThreshold* and *UThreshold* which, depending on the value of *K*, relates to low, medium and high risk in heart data, is calculated the same way as in Section 6.4. Furthermore, the implementation of the scoring mechanism for heart rate in the NEWS algorithm is the same as previous experiments. However, rather than the quartiles approach and NEWS algorithm being assessed independently, the hybrid approach integrates these two decision making systems through a series of conjunction operators. Thus, a pseudocode example of the rule which is created is as follows:

ALERT

IF HeartRate > $UT_{LowRisk}$ AND IF HeartRate <= $UT_{MedRisk}$

AND

IF NEWS_SCORING = TRUE

Through the training process, the quartiles system has prior knowledge on the scopes of normality in a patient's heart data thus no alert is produced if the value is within the bounds of LT and UT. On the other hand, if a heart rate value does fall outside these bounds, the result is passed to the NEWS algorithm and assessed. If the knowledge-based system defines the observed value as normal, i.e. between 51-90 BPM, no action is taken, otherwise an alert is produced. Thus, the above example states that if a heart rate value is considered a low risk and if the NEWS scoring algorithm agrees with this assessment, produce an alert. The quartiles approach aims to act as a false-positive filtering engine, whilst the NEWS algorithm makes use of its predefined scoring system for decision making on a patient's risk score. The implementation of the hybrid approach is achieved using R with source code presented in Appendix D.7.

6.6 Results

To provide ease of reference, the results from all three experiments are presented in Table 6.3 and Table 6.4 for elderly and young patients respectively. The training results, which are applicable in the quartile and hybrid approach, is presented in Table 6.5 and Table 6.6.

6.6.1 NEWS Approach

In the electronic version of the NEWS for risk assessment in heart data, 13 out of 20 tests resulted in an accuracy ratio of 1.00. Of the modelled data that produced false-positives, the most significant result is in the accuracy ratio of Patient's f1o04, f1o09 and f1y04 all of which produced less than 50% accuracy ratio results. Overall, the lowest observed accuracy is 0.14 in the two minute model of Patient f1o09. In other words, nearly all data in this model resulted in the NEWS algorithm raising a false-positive.

Further analysis on each of these patient's real life dataset, i.e. f1o04, f1o09 and f1y04, provides an answer to such a low score. The distribution of all three patient's heart rate, derived from RR intervals, is presented in Figure 6.6.1, and all three examples show that central tendency of values is less than 50 BPM. It should be noted similar results were also found for the real-life data of Patients f1o06 and f1y07 hence the reason these two patients also produced a number of false-positives. In the standard NEWS algorithm for scoring of heart rate, it can be seen that any value between 41 to 50 BPM entails a risk score of 1 (see Table 6.1). Given that the models used to evaluate the NEWS algorithm are based on the first two minutes of the real-life patient's heart data, a wide number of false-positives will be raised for all three patients since their median heart rate all fall under 50 BPM.

The high number of false-positives produced when evaluating the NEWS approach highlights a potential issue with the static nature of this algo-

Table 6.3: Accuracy Ratio Comparison (Elderly Patients)

Dataset	Method	Score of 1	Score of 2	Score of 3	Accuracy Ratio
f1o01	NEWS	0	0	0	1.00
	Quartiles	7	0	0	0.94
	Hybrid	0	0	0	1.00
f1o02	NEWS	0	0	0	1.00
	Quartiles	35	7	0	0.66
	Hybrid	0	0	0	1.00
f1o03	NEWS	0	0	0	1.00
	Quartiles	5	1	2	0.94
	Hybrid	0	0	0	1.00
f1o04	NEWS	66	0	0	0.31
	Quartiles	0	0	0	1.00
	Hybrid	0	0	0	1.00
f1o05	NEWS	0	0	0	1.00
	Quartiles	3	0	0	0.97
	Hybrid	0	0	0	1.00
f1o06	NEWS	34	0	0	0.66
	Quartiles	8	3	0	0.89
	Hybrid	8	0	0	0.92
f1o07	NEWS	0	0	0	1.00
	Quartiles	0	0	0	1.00
	Hybrid	0	0	0	1.00
f1o08	NEWS	0	0	0	1.00
	Quartiles	21	1	0	0.86
	Hybrid	0	0	0	1.00
f1o09	NEWS	63	0	10	0.14
	Quartiles	4	0	0	0.95
	Hybrid	4	0	0	0.95
f1o10	NEWS	0	0	0	1.00
	Quartiles	0	0	0	1.00
	Hybrid	0	0	0	1.00

Table 6.4: Accuracy Ratio Comparison (Young Patients)

Dataset	Method	Score of 1	Score of 2	Score of 3	Accuracy Ratio
f1y01	NEWS	4	0	0	0.97
	Quartiles	2	0	0	0.99
	Hybrid	2	0	0	0.99
f1y02	NEWS	0	0	0	1.00
	Quartiles	13	10	7	0.77
	Hybrid	0	0	0	1.00
f1y03	NEWS	0	0	0	1.00
	Quartiles	5	7	0	0.91
	Hybrid	0	0	0	1.00
f1y04	NEWS	73	0	2	0.17
	Quartiles	3	1	1	0.94
	Hybrid	0	0	0	1.00
f1y05	NEWS	0	0	0	1.00
	Quartiles	2	0	0	0.98
	Hybrid	0	0	0	1.00
f1y06	NEWS	0	0	0	1.00
	Quartiles	4	1	1	0.95
	Hybrid	0	0	0	1.00
f1y07	NEWS	28	0	0	0.74
	Quartiles	0	0	0	1.00
	Hybrid	0	0	0	1.00
f1y08	NEWS	0	0	0	1.00
	Quartiles	7	2	2	0.91
	Hybrid	0	0	0	1.00
f1y09	NEWS	2	0	0	0.99
	Quartiles	7	2	0	0.94
	Hybrid	0	2	0	0.99
f1y10	NEWS	0	0	0	1.00
	Quartiles	0	0	0	1.00
	Hybrid	0	0	0	1.00

Table 6.5: Quartile Training Results (Elderly Patients)

Dataset	Q_1	Q_2	Q_3	$Q_1 - 3(IQR)$	$Q_1 - 2.25(IQR)$	$Q_1 - 1.5(IQR)$	$Q_3 + 1.5(IQR)$	$Q_3 + 2.25(IQR)$	$Q_3 + 3(IQR)$
f1o01	57.25	59.06	60.98	46.08	49.87	51.67	66.56	69.35	72.15
f1o02	55.35	56.39	57.69	48.33	50.08	51.84	61.2	62.97	64.72
f1o03	57.92	59.76	61.48	47.23	49.90	52.57	66.82	69.48	72.16
f1o04	49.02	51.37	54.55	32.44	36.59	40.73	62.83	66.98	71.12
f1o05	53.57	55.35	56.82	43.83	46.27	48.7	61.69	64.12	66.56
f1o06	50.68	51.37	51.9	46.99	47.91	48.83	53.74	54.66	55.59
f1o07	57.03	59.06	61.22	44.46	47.61	50.75	67.51	70.65	73.8
f1o08	68.18	70.42	72.82	54.28	57.76	61.23	79.77	83.24	86.72
f1o09	38.76	40.54	42.25	28.28	30.9	33.52	47.49	50.11	52.74
f1o10	65.22	68.18	71.09	47.6	52	56.41	79.9	84.3	88.71

Table 6.6: Quartile Training Results (Young Patients)

Dataset	Q_1	Q_2	Q_3	$Q_1 - 3(IQR)$	$Q_1 - 2.25(IQR)$	$Q_1 - 1.5(IQR)$	$Q_3 + 1.5(IQR)$	$Q_3 + 2.25(IQR)$	$Q_3 + 3(IQR)$
f1y01	68.18	73.17	78.12	38.35	45.80	53.27	93.04	100.50	107.95
f1y02	54.55	57.25	61.22	34.51	39.51	44.53	71.24	76.25	81.26
f1y03	60.48	63.29	66.37	42.82	47.24	51.65	75.2	79.62	84.04
f1y04	43.1	45.32	47.92	28.64	32.26	35.87	55.15	58.80	62.38
f1y05	53.57	56.82	60.98	31.36	36.91	42.47	72.08	77.64	83.19
f1y06	54.95	58.14	62.5	32.28	37.95	43.61	73.83	79.5	85.16
f1y07	43.73	48.86	56.18	6.39	15.72	25.06	74.85	84.19	93.52
f1y08	56.18	59.29	63.29	34.85	40.18	45.51	73.96	79.29	84.63
f1y09	61.98	65.5	71.09	34.66	41.49	48.32	84.75	91.58	98.41
f1y10	67.87	72.82	77.32	39.53	46.62	53.7	91.49	98.57	105.66

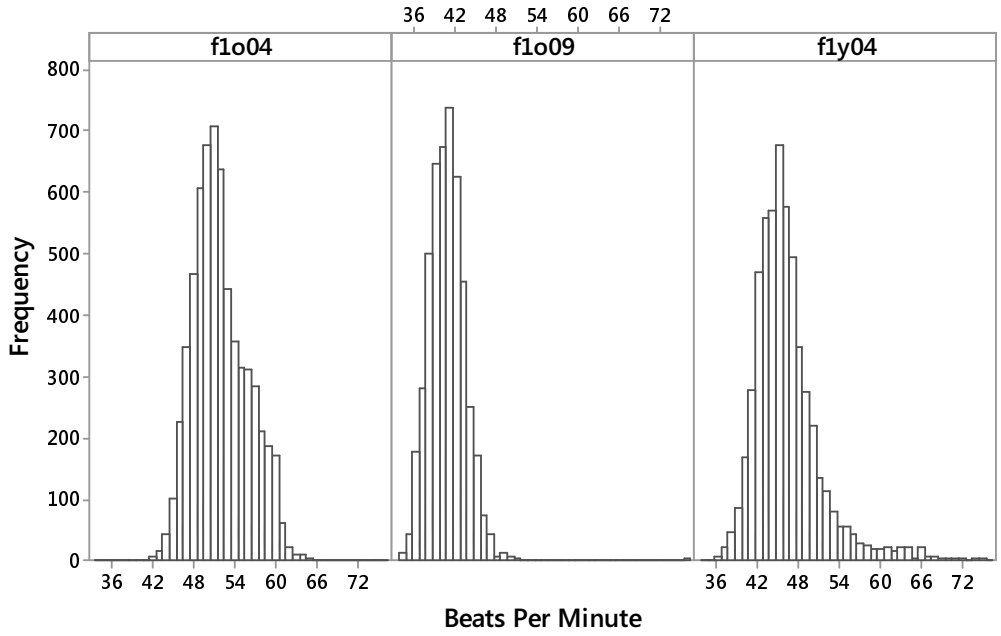

Figure 6.3: Heart Rate Distribution of f1o09 and f1y04

rithm. Because it is more than possible for patients to have a normal healthy heart beat at less than average values, e.g. high-performance athletes, a wide number of false-positives will occur with such patients. This results of this experiment correlates with the observations made in the literature review by [129, 131] in which the argument put forth was that the EWS system is too sensitive resulting in many false-positives when applied in real-life scenarios.

6.6.2 Quartiles Approach

5 out of the 20 tests conducted resulted in an accuracy ratio of 1.00 in the quartile approach. The minimum accuracy ratio observed in this technique is 0.66. The results show that a higher number of false-positives were raised in the quartile approach in comparison with the NEWS algorithm. Such a finding correlates with one of the primary limitations of anomaly detection techniques, as applied to IDS, in which a high number of false-alarms may

be raised as discussed by [132]. Due to the high number of false-positives, one may initially conclude that there is insignificant evidence to demonstrate that the quartiles based approach provides any positive impact to risk assessment of heart data in comparison with the NEWS algorithm. However, the true strength of the quartiles based approach lies in patient's heart data which is outside of normal defined values, namely patients who raised a high number of false-positives using the standard NEWS algorithm.

A primary example is Patient f1o04 where the quartiles based approach produced a minimum accuracy ratio of 0.94 whilst a drastically inferior result is produced by the NEWS algorithm with a ratio of 0.14. Similarly, for Patient's f1o06, f1o09, f1y04 and f1y07 the quartile approach produces less false-positives in comparison with the NEWS algorithm. The wide contrast between results in these patient's demonstrate the primary advantage of the quartile-based approach to defining potential risk in heart data. Given that no predefined static normal heart rate values are classified in the quartiles method, it's proposed and shown that a learning based approach to risk assessment provides greater flexibility for patient's heart data which is out with the normality of the general population.

6.6.3 Hybrid Approach

In the hybrid approach, 16 out of 20 tests resulted in no false-positives being raised. It can be stated that this approach produces less false-positives in comparison with both the NEWS algorithm and the quartile approach. The minimum accuracy ratio for this approach is 0.92 which, once again, is far higher than the previous two techniques.

From the table of results, it can be seen that in most instances of < 1.0 accuracy ratios, the best result from either the NEWS algorithm or quartiles method will be taken by the hybrid approach. This can be explained due to

the conjunction decision making logic whereby both the NEWS algorithm and quartile method must be produce a Boolean statement of true (in terms of risk) in order for a heart rate variable to be considered a risk. From using the hybrid approach, it can be seen that Patient's f1o04, f1o09 and f1y04 produced an accuracy ratio of 1.00, 0.95 and 1.00 which is equal or better than the quartile approach.

Thus, it can be stated that the application of a hybrid-based approach whereby an integration of both expert knowledge and machine-learning - through statistical analysis - provides the ideal approach for reducing the sensitivity in the EWS. However, it is also noted that this approach may potentially be detrimental in detection of actual risks in a patient's heart data, i.e. false negative. Further discussion on the success of this chapter's experiments and shortcomings is noted in the next section prior to conclusion.

6.7 Discussion

Despite the success of the hybrid approach to reduction of false-positives in the EWS system for risk assessment of heart data, there are a few noted limitations in this chapter's work. The first limitation is a practical one which relates to the application of a machine-learning algorithm in real-life scenarios, e.g. hospital environments. The training period applied during the quartiles and hybrid approach consisted of 2 hours of real-life patient data acquired from the Fantasia database. Since the quartiles approach requires a learning window prior to application, one may find that a 2 hour training window is significantly long enough that a patient's health may deteriorate, especially in scenarios where the system is applied to intensive care units. To address the need for a learning window, one may apply the standard static form of the EWS algorithm whilst training occurs before performing a handover to the hybrid approach when sufficient period of training has passed.

The second limitation is that although the hybrid approach performs exceptionally well in suppression of false-positives in normal healthy patient's data, there is of suppression on true-positives when assessing risk heart data. This, in turn, results in false-negatives which is highly detrimental for risk assessment.

Thirdly, a fine balance must be applied to the threshold equations as presented in Section 6.4.2. As noted, the values of K, in the equations of $Q_1 - K(IQR)$ and $Q_3 + K(IQR)$, are derived from the work of Tukey but the author himself states that these values are a *"rule of thumb"* [15]. The limitation in using the standard suggested K values become more transparent when one is to look at the training results for Patient f1y07 in Table 6.6. The calculated values for $Q_1 - 3(IQR)$ and $Q_1 - 2.25(IQR)$ is 6.39 and 15.72. In other words, the quartiles approach for this individual patient states that a medium and high risk is only considered if the heart rate is less than the respective lower thresholds - highly unlikely given that heart rate usually never will be this low in the first place. Thus, scope exists to define values of K which will result in risk thresholds more applicable to human physiology based on expert clinical input

Finally, in reference to physiology, it is noted that this chapter's work has focused specifically on improvements to the sensitivity of the EWS system for specific vital physiological sign. It is proposed that this technique may be applicable to other vital signs, including blood pressure, temperature and respiratory rate. The limitations noted here are addressed in the future work section in the final chapter of this thesis.

6.8 Conclusion

This chapter has provided an evaluation on the accuracy of the NEWS algorithm for assessment of heart data modelled from normal healthy patients. The metric of accuracy has been based on existing work originally

applied to evaluation of intrusion detection systems [13, 14] and is defined as $1 - \frac{FalsePositives}{\sum_{i=1}^{n} s_i}$. The results from simulating one and two minute sample of normal healthy patient's heart data against the NEWS algorithm, in a discrete manner via probabilistic modelling, has shown that this algorithm produces a number of false-positives especially in regards to patient's whose heart rate falls out with the normal ranges of a population. To address the issue of sensitivity in the EWS system, a quartile-based approach, developed originally by [15] and applied to anomaly detection by [16], has shown that this machine-learning algorithm is better capable of learning the ranges of normality from existing patient's heart data through the definition of upper and lower limits.

The results of the quartile-based approach, in comparison with the standard NEWS algorithm, shows that this technique provides better accuracy in patient's whose data falls out with the ranges normally catered for in the EWS algorithms but unfortunately a high number of false-positives are still produced for most data test. Thus, to address the sensitivity of the EWS algorithm along with taking advantage of the quartile-based approach, a hybrid approach is presented which integrates the knowledge-based system with the machine-learning approach to detection of risk in patient's heart data. Results show that the hybrid approach is significantly better than both the NEWS algorithm and quartile approach when applied independently of each other with an 1.00 accuracy for 80% of modelled data and a minimum accuracy ratio of 0.92.

Chapter 7

Conclusion and Future Work

7.1 Thesis Summary

THIS thesis aims to analyse, model and simulate heart data to evaluate a novel Early Warning Score (EWS) algorithm, which enables customisation of risk thresholds for each individual patient, in order to demonstrate increased accuracy on the proposed approach taken for risk assessment.

Heart data, especially heart rate variation (HRV), has been analysed and modelled in past research, but it is identified that there is a disparity of results in regards to the underlying distribution. In Section 3.3, one study conclude that heart data's underlying distribution is a gamma distribution [9] while results from others show normal [8], Weibull [52] and bimodal distributions are also possible candidates [10, 11]. Applied techniques have been proposed for modelling of heart data including the use of Weibull distribution [52], neural networks modelling [103] and normal mixture distribution [102, 58]. In the Weibull distribution method, modelling is achieved using Maximum Likelihood Estimate (MLE) for parameter estimates while mixture distribution makes use of the Expectation Maximization (EM) technique.

In regards to the EWS system, evidence suggests that there is a correlation between the EWS system and predictive power on the deteriorating health of patients [19, 126, 127], however limitations also exist in this system especially the sensitivity of the algorithm as noted by Gao et al. [129]. Johnstone et al. [130] also notes a lack of statistical techniques in the development of such risk assessment systems. From review of Clinical Decision Support Systems (CDSSs) it has been shown that knowledge-based CDSSs, which the electronic EWS falls under, makes use of medical knowledge inputted by an expert whilst machine learning techniques typically use artificial intelligence methods for the decision making process including neural networks and genetic algorithms. Although machine learning techniques have the advantage of demonstrating greater capabilities in the decision making process - without the need for prior knowledge - it is shown that the most commonly used systems are still knowledge-based. The primary reasons for this is the lack of transparency in machine learning implementations [110] resulting in reduced accountability along with a difficulty in such implementations fitting in with the existing work flow of clinical environments [118].

This thesis has argued that no one single distribution is capable of describing and modelling all heart data. Instead, it is proposed that one of four statistical distributions may be used to describe heart data in the general population. Similarly, like heart data, it is argued that the existing EWS algorithms do not provide a viable results in regards to accuracy due to the wide number of variations in the vital signs of a patient. It is proposed that a customisation approach to risk assessment, which considers the normal starting values of patient's parameters which are monitored, produces better accuracy in comparison to a static predefined rule set.

To provide evidence on heart data distribution, the Anderson-Darling (AD) test statistics is used in conjunction with MLE for parameter estimates. Individual models, identified using the AD test, is then compared to an-

other popular technique for modelling of heart data: normal mixture modelling. The approach taken to implementing a customisable EWS algorithm, which caters risk thresholds for each individual patient, is primarily inspired by the anomaly-detection technique originally applied described by Wang et al. [16] whilst calculation of risk thresholds is from the work of Tukey [15]. Two implementations of a customisable EWS algorithm is produced: a quartile-based approach and a hybrid-based approach. To demonstrate viability of the two novel approaches, comparative evaluation against the standard National Early Warning Score (NEWS) algorithm is conducted using modelled data derived from both individual and mixture modelling techniques. An accuracy ratio metric, based upon similar techniques used to evaluate Intrusion Detection Systems (IDSs), is developed which enables quantitative results to be produced as a measurement of false-positives raised by each algorithm in assessing normal healthy patients data.

7.2 Main Findings

The primary findings of this thesis are:

- The underlying distribution of heart data cannot be inferred from one single type of distribution. Thus, Section 4.2.3 hypothesises that such data may fall under one of four distribution types including: normal, logistic, Weibull and gamma distribution. The hypothesised distributions are based upon prior research and findings in the analysis of HRV including the works of [8, 9, 52, 10].

- Through the use of the AD test statistic it is shown that 80% of small samples (up to two minutes) of heart data may be categorised under one of the four hypothesised distributions. The analysis technique applied is similar to [52], however, the use of the AD test in this thesis

has shown that heart data may be described by more than just the Weibull distribution.

- Modelling of heart data is conducted in R ver. 3.0.3, and through a combination of quantitative and qualitative validation results which show that small samples of heart data, namely RR intervals, can be modelled using an individual distribution, when there is significant proof that the data is distributed under the assumed distribution. In the case where the AD test statistic fails, normal mixture modelling, as applied by [102, 58], proves to be valuable in modelling RR intervals when the patients data does not fit a specific distribution or demonstrates bimodal characteristics, i.e. more than one peak in the distribution.

- Through the use of modelled heart data, evaluation results on the three approaches to risk assessment of heart data show the hybrid approach, an integration of both knowledge-based and machine learning techniques, demonstrates the most significant improvement to the accuracy of the EWS system, thus a reduction in the overall number of false-positives when analysing normal healthy patient's heart data.

7.3 Contributions

In achieving the aim of the thesis, the three main contributions made are:

- Statistical analysis and formal hypothesis testing demonstrates that four primary distributions may describe small samples of RR intervals: normal, logistic, Weibull and gamma. RR intervals are the fundamental values of heart data analysed and modelled in this thesis. Qualitative and quantitative validation shows that RR intervals, mod-

elled using one of the four identified distributions, is statistically similar to real-life counterparts. The finding that no one single distribution is capable of describing or modelling each patient's heart data is built upon previous work in the analysis of heart data including [8, 9, 10, 11].

- Recontextualisation of methodologies in evaluation of IDSs, especially the Defense Advanced Research Projects Agency (DARPA) evaluation [12] show that it is possible to apply such techniques to the evaluation of the EWS system in a quantitative manner. Metrics for evaluation in IDS, as originally presented by [13, 14], are modified to produce an accuracy ratio capable of assessing the sensitivity of the EWS algorithm. The accuracy ratio is a measure of false-positives raised when analysing normal healthy patient's heart data.

- Contribution towards improved accuracy in the EWS algorithm produced two novel risk assessment techniques: a quartile-based and hybrid approach. Based on the method defined by [15] and inspired by the work of [16], the quartile approach is an anomaly-based IDS technique which demonstrates a higher degree of accuracy when assessing the risk of healthy patient's heart data which is outside the normal ranges of the general population. A minimum accuracy ratio of 0.14 was observed for the standard EWS algorithm while the quartile-based approach produced a minimum accuracy ratio of 0.66. However, inaccuracies are still produced using the quartile-based technique during some experiments that involved patients with heart rate within the range of normality. Thus, a hybrid approach, integration of both knowledge-based and anomaly-based techniques, demonstrates a higher degree of accuracy in comparison with both the NEWS algorithm and the quartile approach with a minimum accuracy ratio of 0.92 in experiments conducted.

7.4 Future Work

Analysis and modelling of heart data has demonstrated capabilities in describing small samples of RR interval using a single distribution while the use of a quartiles-based approach for risk assessment has produced evidence of greater accuracy in comparison with the standard EWS algorithm. However, several areas for improvement are acknowledged from the experiments conducted. Firstly, analysis and modelling results have shown the capabilities of modelling small samples of heart data (RR intervals) via a discrete probabilistic approach based on MLE parameter estimates and evidence of underlying distribution derived from the AD test statistic. As noted earlier, this method has a success rate of 80%. In the case of non-significant evidence or the need for modelling of heart data greater than approximately two minutes in length, it has been shown that mixture distribution has greater power with respect to this goal.

However, regardless of modelling technique applied, one primary limitation is the lack of accurate simulation of the behaviour in heart data within the time domain. As shown in the discussion of Section 5.4, plotting of modelled heart data as a time series shows that there is a significant difference in comparison with the real-life counterpart. Discussion is given to the concept of the chaotic heart to give justification for this limitation. Proponent and opponent's view on the theory that the heart is a chaotic system is provided in Section 5.4.1 yet fundamental proof for either argument has yet to be formally established. Thus, an area of future work is addressing the question: *is the heart a chaotic system?* In developing a hypothesis and proof to this question, one may be capable of not only producing a more accurate simulation of the heart but also provide greater understanding of this organ from a clinical perspective.

In the scope of risk assessment, results from Chapter 6 has shown that a hybrid approach, using the knowledge of the EWS algorithm and a quar-

tile learning technique, results in a reduced number of false-positives when assessing the risk of a healthy patient's heart data. This approach to risk assessment provides a solution to addressing the issue of sensitivity in existing algorithms as noted by Gao et al. [129]. However, it is acknowledged that the primary limitation of the hybrid approach is the potential for false-negatives to occur, i.e. wrong indication that there is no risk in a patient. Furthermore, the scope of experiments has only been evaluated for one type of vital physiological sign data and it should be reiterated that the EWS algorithm considers multiple parameters including blood pressure, temperature and respiratory rate, as examples. Thus, one area of future work in the usage of a quartile based approach for risk assessment is to evaluate whether this approach works as effectively with other forms of vital sign data. One potential challenge in this task is in the acquisition of additional vital physiological data though this may be partially solved by the Physionet databank.

Definition of suitable K values for the lower and upper threshold calculations, i.e. $Q1 - K(IQR)$ and $Q3 + K(IQR)$, is another area of future work. The K values have been derived from the work of Tukey [15] for this thesis but significant scope and future work exists in this area for further analysis of vital sign data to ensure that K values defined for threshold is clinically reviewed and there is potential to derive values of K catered towards each individual patient. The definition of clinically approved K values may also potentially reduce the number of false-positives raised by the quartile-based approach thus removing the need for a hybrid approach and solving the false-negative implication.

The findings of this thesis, in regards to heart data distribution and techniques in improved accuracy of the EWS system, gives significant potential for future work in clinical risk assessment. In particular, it is proposed that future work can be conducted to bring all findings of this thesis together, along with addressing the limitations of the experiments conducted, to pro-

duce a CDSS which is both accurate, informative and customisable in the goal of defining risk in a patient's vital physiological sign data. Figure 7.1 presents the work flow of a proposed future system which is capable of assessing a patient's well being by observation of both the distributions in vital sign data and the risk score output. Risk thresholds are defined using the quartiles technique; therefore each patient has customised values for increased accuracy while distribution identification of vital sign data can be conducted using the methodology as described in this thesis. In achieving this future goal, it is proposed that not only can the immediate risk of a patient be identified, i.e. using the risk thresholds calculated, but also observation of abnormalities in vital sign data, e.g. the training phase determines that a patient's heart rate is normally distributed, but live monitoring shows that another distribution is observed, which may not pose immediate risk but require additional assessment by a medical expert.

Ultimately, it is recognised that although e-Health, the use of computational power for provision of health care, may provide benefits such as the improving the accuracy of the electronic EWS service, this thesis is in agreement with the recent publication by Carberry et al. [131] that such technology should not replace the expert knowledge of medical practitioners - it should compliment their work.

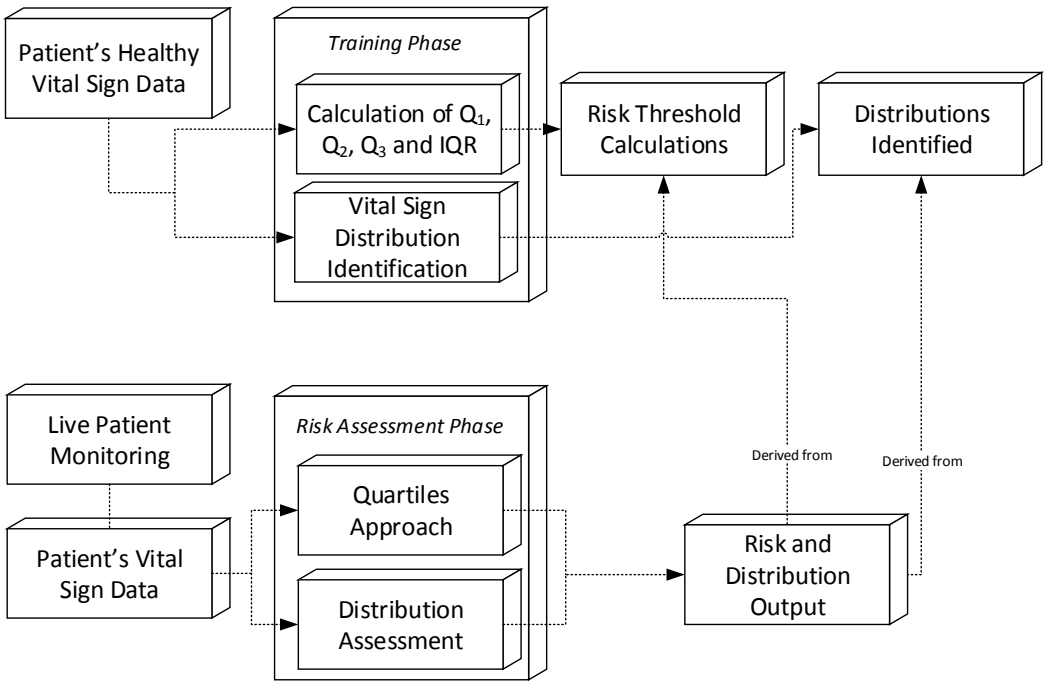

Figure 7.1: Proposed Future Risk Assessment System

References

[1] L. Fan, W. Buchanan, C. Thummler, O. Lo, A. Khedim, O. Uthmani, A. Lawson, and D. Bell. DACAR platform for ehealth services cloud. In *2011 IEEE International Conference on Cloud Computing (CLOUD)*, pages 219–226, 2011.

[2] G. Pesola, H. Pesola, M. J. Nelson, and R. Westfal. The normal difference in bilateral indirect blood pressure recordings in normotensive individuals. *The American Journal of Emergency Medicine*, 19(1):43–45, 2001.

[3] P. Mackowiak, S. Wasserman, and M. Levine. A critical appraisal of 98.6 degrees Fahrenheit, the upper limit of the normal body temperature, and other legacies of carl reinhold august wunderlich. *The Journal of the American Medical Society*, 268(12):1578–1580, 1992.

[4] A. Shoemaker. What's normal? Temperature, gender, and heart rate. *Journal of Statistics Education*, 4(2), 1996.

[5] B. R. O'Driscoll, L. S. Howard, and A. G. Davison. BTS guideline for emergency oxygen use in adult patients. *Thorax*, 63(Suppl 6):vi1–vi68, 2008.

[6] L. Sherwood. *Fundamentals of physiology: a human perspective*. Thomson Brooks/Cole, 2006.

[7] G. Tortora and B. Derrickson. *Principles of anatomy and physiology*. Wiley, 2008.

[8] J. R. Jennings, J. C. Stringfellow, and M. Graham. A comparison of the statistical distributions of beat-by-beat heart rate and heart period. *Psychophysiology*, 11(2):207–210, 1974.

[9] E. Hashida, T. Inoue, N. Yoshitani, and T. Tasaki. A probabilistic study of the irregularity of RR interval durations in atrial fibrillation–a preliminary report. *Japanese circulation journal*, 37(12):1423, 1973.

[10] I. O'Brien, P. O'Hare, and R. Corrall. Heart rate variability in healthy subjects: effect of age and the derivation of normal ranges for tests of autonomic function. *British Heart Journal*, 55(4):348–354, 1986.

[11] J. Tebbenjohanns, B. Schumacher, T. Korte, M. Niehaus, and D. Pfeiffer. Bimodal rr interval distribution in chronic atrial fibrillation. *Journal of cardiovascular electrophysiology*, 11(5):497–503, 2000.

[12] R. Lippmann, J. W. Haines, D. J. Fried, J. Korba, and K. Das. The 1999 DARPA off-line intrusion detection evaluation. *Computer Networks*, 34(4):579–595, 2000.

[13] J. Sommers, V. Yegneswaran, and P. Barford. Toward comprehensive traffic generation for online IDS evaluation. Technical report, University of Wisconsin Technical Report, 2005.

[14] M. G. E. Rab and A. A. E. Kalam. Testing intrusion detection systems: An engineered approach. In *Proceeding of the International Conference on Software Engineering and Applications (SEA 2006)*, 2006.

[15] J. W. Tukey. *Exploratory data analysis*. Addison-Wesley, 1977.

[16] C. Wang, K. Viswanathan, L. Choudur, V. Talwar, W. Satterfield, and K. Schwan. Statistical techniques for online anomaly detection in data centers. In *IFIP/IEEE International Symposium on Integrated Network Management (IM) 2011*, pages 385–392, May 2011.

[17] R. Morgan, F. Williams, and W. MM. An early warning scoring system for detecting developing critical illness. *Clinical Intesive Care*, 8:100, 1997.

[18] U. Kyriacos, J. Jelsma, and S. Jordan. Monitoring vital signs using early warning scoring systems: a review of the literature. *Journal of nursing management*, 19(3):311–330, 2011.

[19] C. Subbe, M. Kruger, P. Rutherford, and L. Gemmel. Validation of a modified early warning score in medical admissions. *QJM: An International Journal of Medicine*, 94(10):521 – 526, 2001.

[20] G. B. Smith, D. R. Prytherch, P. E. Schmidt, and P. I. Featherstone. Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation*, 77(2):170–179, 2008.

[21] D. Goldhill, L. Worthington, A. Mulcahy, M. Tarling, and A. Sumner. The patient-at-risk team: identifying and managing seriously ill ward patients. *Anaesthesia*, 54:853–860, 1999.

[22] A. McGinley and R. M. Pearse. A national early warning score for acutely ill patients. *British Medical Journal*, 345(7869):9, 2012.

[23] P. Bradley. The history of simulation in medical education and possible future directions. *Medical Education*, 40(3):254 – 262, 2006.

[24] R. Zobel. A personal history of simulation in the uk and europe, 1964 - 2001. *Internation Journal of Simulation*, 1(1-2):69–79, 2001.

[25] B. Bernstein. The uneasy alliance: Roosevelt, churchill, and the atomic bomb, 1940-1945. *The Western Political Quarterly*, 29(2):202–230, 1976.

[26] N. Meteopolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

[27] C. A. Murphy and T. Perera. The definition of simulation and its role within an aerospace company. *Simulation Practice and Theory*, 9(6-8):273–291, 2002.

[28] S. Chen and Y. Huang. Risk preference, forecasting accuracy and survival dynamics: Simulations based on a multi-asset agent-based artificial stock market. *Journal of Economic Behavior & Organization*, 67(3-4):702–717, 2008.

[29] A. Maria. Introduction to modeling and simulation. In *Proceedings of the 29th conference on Winter simulation*, pages 7–13. IEEE Computer Society, 1997.

[30] Laerdal. Simman®, 2014. Available at `http://www.laerdal.co.uk/doc/10926039/SimMan.html`.

[31] A. Sulistio, C. S. Yeo, and R. Buyya. A taxonomy of computer-based simulations and its mapping to parallel and distributed systems simulation tools. *Software Practice and Experience*, 34:653–673, June 2004.

[32] J. Banks and J. Carson. *Discrete-Event System Simulation*. Englewood Cliffs, New Jersey: Prentice-Hall, 1984.

[33] I. Mitrani. *Simulation techniques for discrete event systems*. Cambridge University Press, 1982.

[34] B. Jacobs, K. Levy, and H. Markowitz. Financial market simulation. *The Journal of Portfolio Management*, 30(5):142–152, 2004.

[35] NS-2. The network simulator - NS-2, 2011. Available at `www.isi.edu/nsnam/ns/`.

[36] D. Mahrenholz and S. Ivanov. Real-time network emulation with ns-2. In *Eighth IEEE International Symposium on Distributed Simulation and Real-Time Applications, 2004. DS-RT 2004.*, pages 29 – 36, 2004.

[37] L. Muchnik, Y. Louzoun, and S. Solomon. Agent based simulation design principles - applications to stock market. In *Proceedings of the Third Nikkei Econophysics Symposium*, pages 183–188. Springer, 2006.

[38] S. Jamison and M. McCartney. Discrete vs. continuous time implementation of a ring car following model in which overtaking is allowed. *Nonlinear Analysis: Real World Applications*, 10(1):437 – 448, 2009.

[39] J. Schatzmann, K. Georgila, and S. Young. Quantitative evaluation of user simulation techniques for spoken dialogue systems. In *6th SIGdial Workshop on Discourse and Dialogue*, pages 2–3, 2005.

[40] H. Perros. *Computer Simulation Techniques: The definitive introduction!* Online, 2009. Available at `http://www.csc.ncsu.edu/faculty/perros/simulation.pdf`.

[41] W. Gilks, S. Richardson, and D. Spiegelhalter, editors. *Markov Chain Monte Carlo In Practice*. Chapman & Hall, London, 1996.

[42] A. Law. How to conduct a successful simulation study. In *Proceedings of the 2003 Winter Simulation Conference, 2003.*, volume 1, pages 66–70. IEEE, 2004.

[43] K. Liyanage. Methodologies for data collection and model documentation in computer simulation. *International Journal of the Computer, the Internet and Management*, 13(SP2):4.1–4.6, 2005.

[44] W. Trybula. Building simulation models without data. In *IEEE International Conference on Systems, Man, and Cybernetics, 1994. Humans, Information and Technology.*, volume 1, pages 209–214, Oct 1994.

[45] N. Robertson and T. Perera. Automated data collection for simulation? *Simulation Practice and Theory*, 9(6-8):349–364, 2002.

[46] R. Sargent. Verification and validation of simulation models. In *Proceedings of the 40th Conference on Winter Simulation*, WSC '08, pages 157–169. Winter Simulation Conference, 2008.

[47] A. Law and W. Kelton. *Simulation Modeling and Analysis*, volume 2. McGraw-Hill New York, 1991.

[48] D. C. Montgomery and G. C. Runger. *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.

[49] B. Martin. *Statistics for Physicists*. Academic Press, 1971.

[50] F. Owen and R. Jones. *Statistics*. Polytech Publishers, 1977.

[51] D. Montgomery and G. Runger. *Applied Statistics and Probability for Engineers*. Wiley-India, 4 edition, 2007.

[52] S. J. Mandrekar, H. N. Nagaraja, and G. G. Berntson. Statistical modelling of the differences between successive R–R intervals. *Statistics in medicine*, 24(3):437–451, 2005.

[53] Madhero88. File:ECGcolor.svg. Licensed under Public domain via Wikimedia Commons. Available at `http://commons.wikimedia.org/wiki/File:ECGcolor.svg`.

[54] E. Roth. 24-hour holter monitoring, August 2012. Available at `http://www.healthline.com/health/holter-monitor-24h`.

[55] Tristaess. File:btl-holter p-unit-4-leads 0608.jpg, 2010. Licensed under Public domain via Wikimedia Commons. Available at `http://commons.wikimedia.org/wiki/File:BTL-holter_P-unit-4-leads_0608.jpg`.

[56] G. G. Berntson. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34:623–648, 1997.

[57] G. E. Billman. Heart rate variability–a historical perspective. *Frontiers in physiology*, 2:1–13, 2011.

[58] J. M. Ketchum, A. M. Best, and V. Ramakrishnan. A within-subject normal-mixture model with mixed-effects for analyzing heart rate variability. *Journal of Biometrics & Biostatistics*, s7:1–6, 2012.

[59] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Available at: `http://circ.ahajournals.org/cgi/content/full/101/23/e215`.

[60] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, 1996.

[61] G. Pinna, R. Maestri, A. Di Cesare, R. Colombo, and G. Minuco. The accuracy of power-spectrum analysis of heart-rate variability from annotated RR lists generated by holter systems. *Physiological measurement*, 15(2):163–179, 1994.

[62] R. Hesselfeldt, M. Kristensen, and L. Rasmussen. Evaluation of the airway of the SimMan® full-scale patient simulator. *Acta Anaesthesiologica Scandinavica*, 49(9):1339–1345, 2005.

[63] A. Wyatt, F. Archer, and B. Fallows. Use of simulators in teaching and learning: Paramedics' evaluation of a patient simulator. *Journal of Emergency Primary Health Care*, 5:1–16, 2007.

[64] Y. Hwang, S. Lampotang, N. Gravenstein, I. Luria, and B. Lok. Integrating conversational virtual humans and mannequin patient simulators to present mixed reality clinical training experiences. In *8th*

IEEE International Symposium on Mixed and Augmented Reality. ISMAR 2009., pages 197–198, 2009.

[65] V. Howard, C. Ross, A. Mitchell, and G. Nelson. Human patient simulators and interactive case studies. *CIN: Computers, Informatics, Nursing*, 28:42–48, 2010.

[66] L. Meng and T. Spedding. Modelling patient arrivals when simulating an accident and emergency unit. In *Proceedings of the 2008 Winter Simulation Conference, 2008.*, pages 1509–1515, 2008.

[67] E. Kolb, J. Peck, S. Schoening, and T. Lee. Reducing emergency department overcrowding - five patient buffer concepts in comparison. In *Proceedings of the 2008 Winter Simulation Conference*, pages 1516–1525, 2008.

[68] O. Balci. Principles and techniques of simulation validation, verification, and testing. In *Proceedings of the 1995 Winter Simulation, 1995.*, pages 147–154, 1995.

[69] B. Agar, M. Eren, and A. Cinar. Glucosim: educational software for virtual experiments with patients with type 1 diabetes. In *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005.*, pages 845–848. IEEE, 2006.

[70] D. S. Eng and J. M. Lee. The promise and peril of mobile health applications for diabetes and endocrinology. *Pediatric diabetes*, 14(4):231–238, 2013.

[71] T. E. Brown, L. A. Beightol, J. Koh, and D. L. Eckberg. Important influence of respiration on human RR interval power spectra is largely ignored. *Journal of Applied Physiology*, 75(5):2310–2317, 1993.

[72] M. P. Tulppo, T. Makikallio, T. Takala, T. Seppanen, and H. Huikuri. Quantitative beat-to-beat analysis of heart rate dynamics during ex-

ercise. *American Journal of Physiology-Heart and Circulatory Physiology*, 271(1):H244–H252, 1996.

[73] J. A. Taylor and D. L. Eckberg. Fundamental relations between short-term RR interval and arterial pressure oscillations in humans. *Circulation*, 93(8):1527–1532, 1996.

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

[75] M. Wolf, G. Varigos, D. Hunt, and J. Sloman. Sinus arrhythmia in acute myocardial infarction. *The Medical Journal of Australia*, 2(2):52–53, 1978.

[76] H. L. Leistner, G. G. Haddad, R. A. Epstein, T. L. Lai, M. A. F. Epstein, and R. B. Mellins. Heart rate and heart rate variability during sleep in aborted sudden infant death syndrome. *The Journal of Pediatrics*, 97(1):51–55, 1980.

[77] F. Perticone, R. Ceravolo, R. Maio, C. Cosco, and P. L. Mattioli. Heart rate variability and sudden infant death syndrome. *Pacing and Clinical Electrophysiology*, 13(12):2096–2099, 1990.

[78] K. Antila, I. Vlimki, M. Mkel, J. Tuominen, A. Wilson, and D. Southall. Heart rate variability in infants subsequently suffering sudden infant death syndrome (SIDS). *Early Human Development*, 22(2):57–72, 1990.

[79] S. K. Mehta, D. M. Super, D. Connuck, A. Salvator, L. Singer, L. G. Fradley, R. A. Harcar-Sevcik, H. L. Kirchner, and E. S. Kaufman. Heart rate variability in healthy newborn infants. *The American journal of cardiology*, 89(1):50–53, 2002.

[80] A. Law. A tutorial on how to select simulation input probability distributions. In *Proceedings of the Winter Simulation Conference*, WSC '12, pages 165:1–165:15. Winter Simulation Conference, 2012.

[81] V. Ricci. Fitting distributions with R. *Contributed Documentation available on CRAN*, 96:1–24, 2005.

[82] K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894.

[83] J. B. Ramsey. *The Elements of Statistics with Applications to Economics and the Social Sciences*. Cengage Learning, 2001.

[84] J. Bickel and A. Doksum. *Mathematical Statistics Basic Ideas and Selected Topics*, volume 1. Prentice Hall, 2001.

[85] J. Aldrich. R.a. fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3):162–176, 1997.

[86] I. J. Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100, 2003.

[87] G. Casella and R. L. Berger. *Statistical inference*. Duxbury Press Belmont, CA, 2 edition, 2002.

[88] D. P. Loucks, E. Van Beek, J. R. Stedinger, J. P. Dijkman, and M. T. Villars. *Water resources systems planning and management: an introduction to methods, models and applications*. Paris: UNESCO, 2005.

[89] D. Bilkova. Estimating parameters of lognormal distribution using the method of l-moments. *Research Journal of Economics, Business and ICT*, 4:4–9, 2012.

[90] M. Akram and A. Hayat. Comparison of estimators of the weibull distribution. *Journal of Statistical Theory and Practice*, 8(just-accepted):238–259, 2014.

[91] T. W. Anderson and D. A. Darling. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 23:193–212, 1952.

[92] M. Stephens. The goodness-of-fit statistic v_n: Distribution and significance points. *Biometrika*, 52:309–321, 1965.

[93] M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737, 1974.

[94] M. A. Stephens. Asymptotic results for goodness-of-fit statistics with unknown parameters. *The Annals of Statistics*, 4:357–369, 1976.

[95] M. A. Stephens. Goodness of fit for the extreme value distribution. *Biometrika*, 64(3):583–588, 1977.

[96] M. A. Stephens. Goodness of fit tests with special reference to tests for exponentiality. Technical report, DTIC Document, 1978.

[97] M. Stephens. Tests of fit for the logistic distribution based on the empirical distribution function. *Biometrika*, 66(3):591–595, 1979.

[98] Z. Drezner, O. Turel, and D. Zerom. A modified kolmogorov–smirnov test for normality. *Communications in Statistics-Simulation and Computation®*, 39(4):693–704, 2010.

[99] N. M. Razali and Y. B. Wah. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1):21–33, 2011.

[100] J. Cohen. Statistical power analysis. *Current directions in psychological science*, 1(3):98–101, 1992.

[101] S. S. Shapiro and M. B. Wilk. *An analysis of variance test for normality(complete samples).* PhD thesis, JSTOR, 1964.

[102] T. Costa, G. Boccignone, and M. Ferraro. Gaussian mixture model of heart rate variability. *PloS one*, 7(5):1–9, 2012.

[103] S. Saalasti. *Neural networks for heart rate time series analysis*. Jyväskylän yliopisto, 2003.

[104] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[105] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, and P. C. Tang. Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association*, 8(6):527–534, 2001.

[106] A. Berlin, M. Sorani, and I. Sim. A taxonomic description of computer-based clinical decision support systems. *Journal of Biomedical Informatics*, 39(6):656–667, 2006.

[107] M. Marcos, J. A. Maldonado, B. Martnez-Salvador, D. Bosc, and M. Robles. Interoperability of clinical decision-support systems and electronic health records using archetypes: A case study in clinical trial eligibility. *Journal of Biomedical Informatics*, 46(4):676–689, 2013.

[108] S. H. El-Sappagh and S. El-Masri. A distributed clinical decision support system architecture. *Journal of King Saud University - Computer and Information Sciences*, 26(1):69–78, 2014.

[109] M. A. Musen, Y. Shahar, and E. Shortliffe. Clinical decision-support systems. In *Biomedical Informatics*, Health Informatics, pages 698–736. Springer New York, 2006.

[110] E. S. Berner and T. La Lande. Overview of clinical decision support systems. In *Clinical Decision Support Systems Theory and Practice*. Springer, 2007.

[111] B. G. Buchanan and E. H. Shortliffe. *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project (The Addison-Wesley series in artificial intelligence)*. Addison-Wesley Longman Publishing Co., Inc., 1984.

[112] S. L. Achour, M. Dojat, C. Rieux, P. Bierling, and E. Lepage. A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development. *Journal of the American Medical Informatics Association*, 8(4):351–360, 2001.

[113] S. Rahaman and M. Hossain. A belief rule based clinical decision support system to assess suspicion of heart failure from signs, symptoms and risk factors. In *International Conference on Informatics, Electronics Vision (ICIEV), 2013*, pages 1–6, May 2013.

[114] E. Al-Shaer and H. Hamed. Firewall policy advisor for anomaly discovery and rule editing. In *IFIP/IEEE Eighth International Symposium on Integrated Network Management, 2003.*, pages 17–30, March 2003.

[115] P. J. Lisboa and A. F. Taktak. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural networks*, 19(4):408–415, 2006.

[116] M. Narasingarao, R. Manda, G. Sridhar, K. Madhu, and A. Rao. A clinical decision support system using multilayer perceptron neural network to assess well being in diabetes. *Journal of the Association of Physicians of India*, 57:127–133, 2009.

[117] A. Mangalampalli, S. M. Mangalampalli, R. Chakravarthy, and A. K. Jain. A neural network based clinical decision-support system for efficient diagnosis and fuzzy-based prescription of gynecological diseases using homoeopathic medicinal system. *Expert Systems with Applications*, 30(1):109–116, 2006. Intelligent Bioinformatics Systems.

[118] P. J. Lisboa. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural networks*, 15(1):11–39, 2002.

[119] D. Beasley, R. Martin, and D. Bull. An overview of genetic algorithms: Part 1. fundamentals. *University computing*, 15:58–58, 1993.

[120] B. B. Zellner, S. D. Rand, R. Prost, H. Krouwer, and V. Chetty. A cost-minimizing diagnostic methodology for discrimination between neoplastic and non-neoplastic brain lesions: utilizing a genetic algorithm. *Academic Radiology*, 11(2):169–177, 2004.

[121] P. P. S. Tomar and R. Singh. Evolutionary continuous genetic algorithm for clinical decision support system. *African Journal of Computing & ICT*, 6(1):127–140, 2013.

[122] S. Amin, K. Agarwal, and R. Beg. Genetic neural network based data mining in prediction of heart disease using risk factors. In *2013 IEEE Conference on Information Communication Technologies (ICT)*, pages 1227–1231, April 2013.

[123] F. Sufi, Q. Fang, and I. Cosic. A mobile phone based intelligent scoring approach for assessment of critical illness. In *International Conference on Information Technology and Applications in Biomedicine, 2008. ITAB 2008.*, pages 290 –293, 2008.

[124] D. de Jager, E. Mazomenos, A. Banerjee, K. Maharatna, and J. Reeve. A low-power simplified-MEWS scoring device for patient monitoring. In *4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010.*, pages 1–4, 2010.

[125] T. O'Kane, J. O'Donoghue, J. Gallagher, A. Aftab, A. Casey, and G. Courtney. Electronic-early warning scorecard: An intelligent context aware decision making approach for patient monitoring. In

Handbook of Ambient Assisted Living, volume 11. IOS Press Ebooks, 2012.

[126] J. Gardner-Thorpe, N. Love, J. Wrightson, S. Walsh, and N. Keeling. The value of modified early warning score (MEWS) in surgical in-patients: a prospective observational study. *Annals of the Royal College of Surgeons of England*, 88(6):571–575, 2006.

[127] D. R. Goldhill, A. F. McNarry, G. Mandersloot, and A. McGinley. A physiologically-based early warning score for ward patients: the association between score and outcome. *Anaesthesia*, 60(6):547–553, 2005.

[128] K. Hillman, J. Chen, M. Cretikos, R. Bellomo, D. Brown, G. Doig, S. Finfer, A. Flabouris, M. S. Investigators, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet*, 365(9477):2091–2097, 2005.

[129] H. Gao, A. McDonnell, D. Harrison, T. Moore, S. Adam, K. Daly, L. Esmonde, D. Goldhill, G. Parry, A. Rashidian, C. Subbe, and S. Harvey. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Medicine*, 33(4):667–679, 2007.

[130] C. Johnstone, J. Rattray, and L. Myers. Physiological risk factors, early warning scoring systems and organizational changes. *Nursing in Critical Care*, 12(5):219–224, 2007.

[131] M. Carberry, P. Clements, and E. Headley. How helpful are early warning scores? *Nursing times*, 110(1-3):12–14, 2014.

[132] H. Debar, M. Dacier, and A. Wespi. Towards a taxonomy of intrusion detection systems. *Computer Network*, 31(9):805–822, 1999.

[133] M. Roesch. Snort: Lightweight intrusion detection for networks. In *Proceedings of LISA '99: 13th Systems Administration Conference*, volume 99, pages 229–238, 1999.

[134] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez. Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 28(1-2):18–28, 2009.

[135] D. E. Denning. An intrusion-detection model. *IEEE Transactions on Software Engineering*, (2):222–232, 1987.

[136] P. Helman and G. Liepins. Statistical foundations of audit trail analysis for the detection of computer misuse. *IEEE Transactions on Software Engineering*, 19(9):886–901, Sep 1993.

[137] M. J. Ranum. Experiences benchmarking intrusion detection systems. Technical report, NFR Security White Paper, 2001.

[138] L. M. Rossey, R. K. Cunningham, D. J. Fried, J. C. Rabek, R. P. Lippmann, J. W. Haines, and M. A. Zissman. Lariat: Lincoln adaptable real-time information assurance testbed. In *Aerospace Conference Proceedings, 2002. IEEE*, volume 6, pages 6–2671. IEEE, 2002.

[139] F. Massicotte, F. Gagnon, Y. Labiche, L. Briand, and M. Couture. Automatic evaluation of intrusion detection systems. In *22nd Annual Computer Security Applications Conference, 2006. ACSAC '06.*, pages 361–370, Dec 2006.

[140] N. Athanasiades, R. Abler, J. Levine, H. Owen, and G. Riley. Intrusion detection testing and benchmarking methodologies. In *Proceedings of the First IEEE International Workshop on Information Assurance, 2003. IWIAS 2003.*, pages 63–72. IEEE, 2003.

[141] C. Thomas, V. Sharma, and N. Balakrishnan. Usefulness of DARPA dataset for intrusion detection system evaluation. In *SPIE Defense and Security Symposium*, pages 69730G–69730G. International Society for Optics and Photonics, 2008.

[142] M. L. Shyu, S. C. Chen, K. Sarinnapakorn, and L. Chang. A novel anomaly detection scheme based on principal component classifier. Technical report, University of Miami Department of Electrical and Computer Engineering, 2003.

[143] Lincoln Laboratory. DARPA intrusion detection data sets, 2014. Available at `http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/`.

[144] Information and Computer Science. KDD cup 1999 data, 1999. Available at `http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html`.

[145] M. Tavallaee, N. Stakhanova, and A. Ghorbani. Toward credible evaluation of anomaly-based intrusion-detection methods. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(5):516–524, Sept 2010.

[146] B. Caswell, J. Beale, and A. Baker. *Snort Intrusion Detection and Prevention Toolkit*. Syngress, 2007.

[147] N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 40(4):R1078, 1996.

[148] PhysioNet. The WFDB software package, December 2013. Available at `http://www.physionet.org/physiotools/wfdb.shtml`.

[149] PhysioNet. Physiobank annotations, March 2013. Available at `http://www.physionet.org/physiobank/annotations.shtml`.

[150] A. J. Moss. Measurement of the QT interval and the risk associated with QTc interval prolongation: A review. *The American Journal of Cardiology*, 72(6):23B–25B, 1993.

[151] I. Goldenberg, A. J. Moss, W. Zareba, et al. QT interval: how to measure it and what is" normal". *Journal of cardiovascular electrophysiology*, 17(3):333–336, 2006.

[152] Minitab. Minitab statistical software, release 17 for windows, 2014. Available at `http://www.minitab.com/en-us/`.

[153] F. Lombardi. Chaos theory, heart rate variability, and arrhythmic mortality. *Circulation*, 101(1):8–10, 2000.

[154] J. Knox. The effect of postural and exercise components on the heart rate during a brief step test. *The Journal of physiology*, 108(3):340–344, 1949.

[155] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[156] A. P. Dempster, N. M. Laird, D. B. Rubin, et al. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal statistical Society*, 39(1):1–38, 1977.

[157] J. A. Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.

[158] T. L. Bailey and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Intelligent Systems for Molecular Biology*, 1994.

[159] D. McLernon, N. Dabanloo, A. Ayatollahi, V. Majd, and H. Zhang. A new nonlinear model for generating rr tachograms. In *Computers in Cardiology, 2004*, pages 481–484, Sept 2004.

[160] G. Georgieva-Tsaneva, M. Gospodinov, and E. Gospodinova. Simulation of heart rate variability data with methods of wavelet transform. In *Proceedings of the 13th International Conference on Computer Systems and Technologies*, CompSysTech '12, pages 306–312, New York, NY, USA, 2012. ACM.

[161] L. A. Lipsitz and A. L. Goldberger. Loss of "complexity" and aging: potential applications of fractals and chaos theory to senescence. *Jama*, 267(13):1806–1809, 1992.

[162] A. L. Goldberger. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *The Lancet*, 347(9011):1312–1314, 1996.

[163] G. Krstacic, D. Gamberger, A. Krstacic, T. Smuc, and D. Milicic. The chaos theory and non-linear dynamics in heart rate variability in patients with heart failure. In *Computers in Cardiology, 2008*, pages 957–959, Sept 2008.

[164] J. K. Kanters, N.-H. Holstein-Rathlou, and E. Agner. Lack of evidence for low-dimensional chaos in heart rate variability. *Journal of cardiovascular electrophysiology*, 5(7):591–601, 1994.

[165] L. Glass. Introduction to controversial topics in nonlinear science: Is the normal heart rate chaotic? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(2):1–4, 2009.

[166] R. Femat, J. Alvarez-Ramirez, and M. Zarazua. Chaotic behavior from a human biological signal. *Physics Letters A*, 214(3-4):175–179, 1996.

[167] D. S. Coffey. Self-organization, complexity and chaos: the new biology for medicine. *Nature medicine*, 4(8):882–885, 1998.

[168] D. Avnir, O. Biham, D. Lidar, and O. Malcai. Is the geometry of nature fractal? *Science*, 279(5347):39–40, 1998.

[169] O. Lo, J. Graves, and W. J. Buchanan. Towards a framework for the generation of enhanced attack/background network traffic for evaluation of network-based intrusion detection systems. In *European Conference on Information Warefare 2010*, 2010.

[170] S. Nirenburg, M. McShane, S. Beale, T. P. O'Hara, B. Jarrell, G. Fantry, and J. Raczek. Cognitive simulation in virtual patients. In *FLAIRS Conference*, pages 174–175, 2006.

[171] M. M. Bouamrane, A. Rector, and M. Hurrell. Using ontologies for an intelligent patient modelling, adaptation and management system. In R. Meersman and Z. Tari, editors, *On the Move to Meaningful Internet Systems: OTM 2008*, volume 5332 of *Lecture Notes in Computer Science*, pages 1458–1470. Springer Berlin Heidelberg, 2008.

[172] S. Nirenburg, M. McShane, S. Beale, B. Jarrell, and G. Fantry. Integrating cognitive simulation into the maryland virtual patient. *Studies in Health Technologies and Informatics*, 142:224 – 229, 2009.

[173] C. Rolim, F. Koch, C. Westphall, J. Werner, A. Fracalossi, and G. Salvador. A cloud computing solution for patient's data collection in health care institutions. In *Second International Conference on eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED '10.*, pages 95 –99. IEEE, 10-16 2010.

[174] S. Troschuetz. .NET random number generators and distributions, May 2007. Available at `http://www.codeproject.com/KB/recipes/Random.aspx`.

[175] O. Lo, L. Fan, B. W., and C. Thuemmler. Conducting performance evaluation of an e-health platform. In *Information Systems and Technology for Organizations in a Networked Society*, chapter 16, pages 295–315. Society IGI Global Publishing, 2013.

[176] M. Edmead and P. Hinsberg. Performance monitor counters, 2011. Available at `http://technet.microsoft.com/en-us/library/cc768048.aspx`.

[177] Microsoft. Stopwatch class, 2011. Available at `http://msdn.microsoft.com/en-us/library/system.diagnostics.stopwatch.aspx` [Accessed 14 November 2011].

[178] M. Franklin and S. Zdonik. Data in your face: push technology in perspective. *SIGMOD Rec.*, 27:516–519, June 1998.

[179] M. B. Smith. An introduction to the windows time service, January 2010. Available at `http://tinyurl.com/c22b53n` [Accessed 1 February, 2012].

Appendix A

Patient Simulator and Cloud4Health Evaluation

PRIOR to refinement in scope of aim and objectives, work was conducted in implementation of a patient simulator designed for simulation of all five key vital physiological signs including heart rate, blood pressure, temperature, oxygen levels and respiratory rate. The design of this simulator along is presented in this chapter. The primary goal of the simulator was to provide demonstration of the applicability of the Cloud4Health project as discussed in the introduction of this thesis. Evaluation results of the Cloud4Health project, via the use of the patient simulator, is discussed from Section A.3 onwards.

It should be noted the work presented in this section was conducted prior to further analysis on heart data, thus although the methodology applied for modelling of vital sign data is valid, i.e. the use of normal distribution only, findings from the core text of this thesis shows that there is future scope in the improvement to modelling techniques applied to the patient simulator proposed.

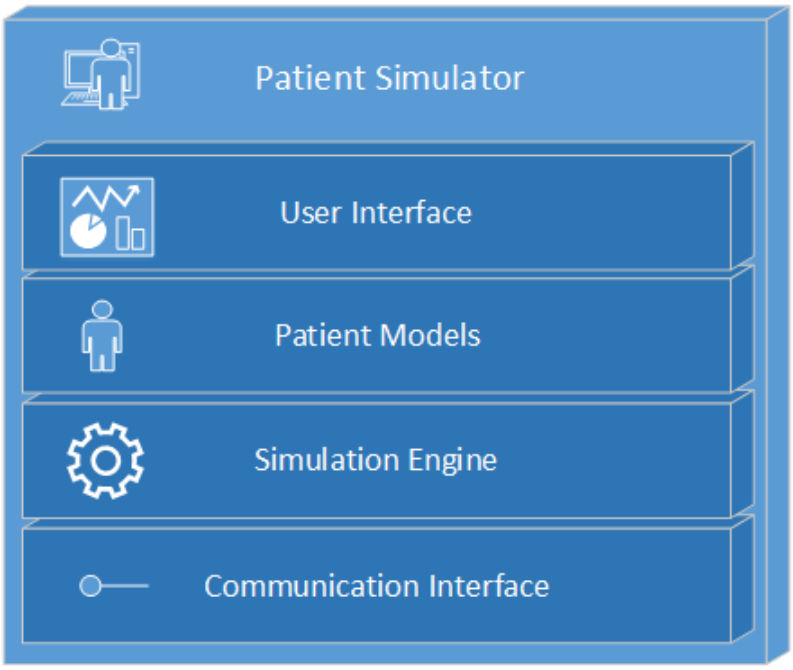

Figure A.1: Patient Simulation Framework

A.1 Patient Simulator Framework

A four tiered approach has been taken in defining the patient simulator framework. The four components which make up the simulator include: user interface, patient models, simulation engine and communication engine. An abstract view of the framework is presented in Figure A.1. A discussion on each of the four components of this framework is presented in this section.

A.1.1 User Interface

As highlighted by [31], the user interface of a simulation provides a far easier means of running a simulator in comparison with a command line driven interface. It is acknowledged that this part of the framework is not an essential necessity in conducting research. However, as demonstrated in previous research [169], a user interface can bring about ease-of-use, automation and simplicity in conducting testing and validation tasks. In other words, it makes running the application a lot simpler.

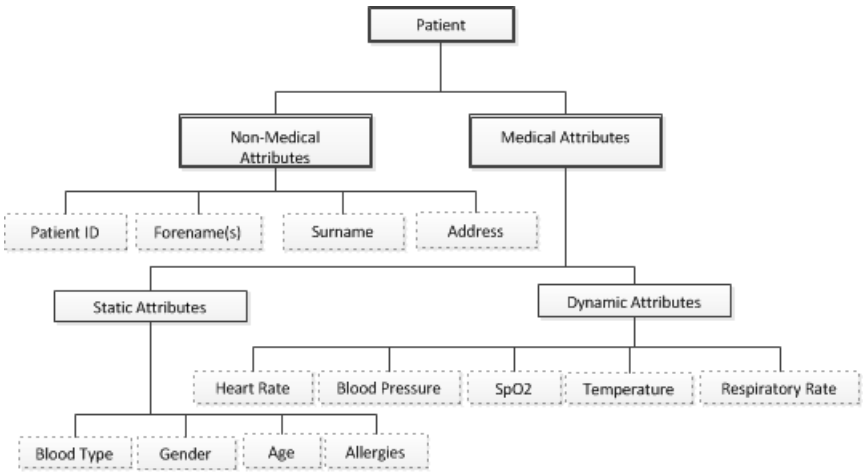

Figure A.2: Patient Model

A.1.2 Patient Models

A variety of ontologies have been proposed by researchers in defining a model of a patient including [170, 171, 172]. The Patient Simulator adopts a simplified ontology, whereby two main categories of attributes are considered: non-medical attributes and medical attributes. Within the subclass of medical attributes, these may be further split into dynamic medical attributes and static medical attributes. Figure A.2 shows the current model of a patient.

Non-medical attributes refer to attributes which, though important, do not have significance when applied to a health care environment. In other words, medical staff will not take these attributes into consideration when it comes to diagnosing a patient's health. Examples of non-medical attributes are:

- *Blood Type* -Blood type of patient (i.e. A, B, AB or O)
- *Gender* - Male or Female
- *DoB* - Date of Birth of a patient

Finally, dynamic medical attributes relate to a patient's vital physiological signs, which have the characteristics of discrete change throughout a

patient's stay in a health care facility. Examples of the dynamic medical attributes include:

- *HR* - The heart rate of a patient, measured in the unit of beats per minute (BPM)
- *BP* - The systolic blood pressure of a patient, measured in mmHg (millimetres of mercury)
- *Temp* - Temperature of a patient's body, measured in degree Celsius (°C)
- *SpO2* - Oxygen level of a patient, measured in percentage (%)
- *RR* - Respiratory rate of patient, measured in Breathing Frequency (BF) per minute

The dynamic medical attributes which have been chosen to simulate in a patient is based upon the concept of the Early Warning Score (EWS) system discussed in prior chapters of this thesis. To provide a brief reiteration, the EWS was originally developed by Morgan et al. [17] and is a risk based scoring system used by nurses and other healthcare staff in rating a patient's health status (e.g. in Accident and Emergency departments). In essence, a higher score suggests a greater risk to a patient's life. Traditionally, the EWS system has been applied as a paper-based observation chart whereby the risk is calculated manually by a member of hospital staff.

Several attributes are taken into consideration when calculating the risk of a patient. Generally, vital physiological sign parameters including heart rate, blood pressure, body temperature, respiratory rate and oxygen levels are observed [19]. Medical staff who carry out observation of the patient will assign a numeric value for each parameter and all values calculated together gives the patient's risk score. By modelling the dynamic medical attributes on the vital physiological signs which are observed in the EWS

system, this allows for the patient simulator to cover one of the key group of attributes of patient data.

A.1.3 Simulation Engine

The simulation engine forms the core aspect of the patient simulator framework. It is within this component that the actual simulation process takes place, including simulation of both medical and non-medical attributes of a patient. In this framework, it is proposed that the use of Discrete-Event Based Simulation (DES) along with mixture of both probabilistic and deterministic behaviour. The simulation engine will manage the time intervals during simulation along with producing output values during simulation.

To justify the choice in DES, it should be reiterated at this point that the goal of this research is in the simulation of patient data rather than simulation of the complete human body. Continuous-Event Based Simulation (CES) would be the ideal choice for simulation of the human body since this system works in a continuous manner but for patient data interactions only occur at certain time intervals. To put this into perspective, consider interactions in a real-life healthcare environment. In the case of a nurse taking a temperature of a patient, this task will only ever be conducted at certain time intervals. Similar to manual methods, in using e-Health services, the temperatures of a real-life patient will only be read at certain time intervals before being uploaded to a storage system. Hence, using DES method allows for the simulation of such interactions with patient data as would be seen in real life.

As part of the simulation engine, a choice of both probabilistic and deterministic behaviour outputs has been chosen. In the simulation of static and non-medical attributes, deterministic behaviour of output is simulated. However, in the case of dynamic medical attributes (vital physiological signs), probabilistic behaviour is chosen. Vital physiological signs are, nat-

urally, dynamic hence there is a degree of variability in a patients vital signs at any time interval. Therefore, using probabilistic behaviour, in the simulation of vital signs, enables more realistic simulation of these attributes. Section A.1.5 discusses the specific technique applied for simulation of vital sign data.

A.1.4 Communication Interface

The final component of the propose patient simulator framework is the communication interface. It is proposed this part of the framework enables communication with e-Health infrastructures using standardised protocols. Examples of such protocols include XML, HTTP and WML as used in the e-Health implementation by [173] and SOAP based protocol as used by [1].

The choice of protocol used will be entirely dependent on the supported communication method of the e-Health environment under evaluation. By defining the communication interface as a separate component, future implementations of the patient simulator can be easily upgraded and extended without other components, i.e. simulation engine and patient models, being modified. In other words, the communication interface enables interoperability with both existing, and future, e-Health environments.

A.1.5 Generation of Vital Physiological Signs

The probability statistics concept of normal distribution has been applied for the generation of vital physiological sign values. Alternatively known as Gaussian Distribution [49], normal distribution is the theory that by generating a set of random values, and applying the mean (average) and standard deviation (variance), the values will tend to cluster around the mean [50, 51].

Defining a mean value for each vital physiological sign is outlined later

on in this section but, unfortunately, determining the standard deviation proves to be a lot more difficult. In determining the standard deviation which would be applied to heart rate, as an example, one must consider the heart rate variation (HRV) of a patient. As discussed in the primary text of this thesis, Section 5.4, difficulties were faced in realistically modelling the subtle behaviour of heart variation. Given complexities in defining variation for heart rate (and the other four vital physiological signs), by default, the Patient Simulator applies a arbitrary standard deviation for each of the vital signs.

Table A.1 presents the default mean and arbitrary standard deviation applied to all five vital physiological signs which are simulated in a patient by default. The mean value of blood pressure is based on the work of Pesola et al. in which they state that a normal systolic blood pressure is found to be 112 mmHg [2]. From studies carried out by both Mackowiak et al. and Shoemaker the result of 36.8 °C [3, 4] is applied for the mean body temperature. O'Driscoll et al. defines normal Spo2 as 96-98% [5], hence the average value of 97% is used. Finally, both Sherwood and Tortora et al. agree that the mean respiration rate is found to be 12 breaths per minute [6, 7].

Table A.1: Default Mean and Standard Deviation of Vital Signs

Vital Sign	**Mean**	**Standard Deviation**
Heart Rate	80 BPM	1.5
Blood Pressure	112 mmHg	2
Body Temperature	36.8°C	0.2
SpO2	97%	0.5
Respiration Rate	12 breaths per min	1

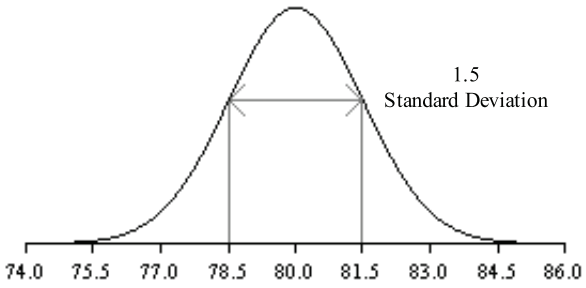

Figure A.3: Normal Distribution of Heart Rate

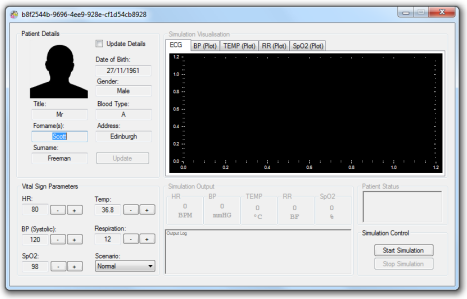

Figure A.4: Patient Simulator GUI

A.2 Implementation of Simulator

The Patient Simulator was implemented using Microsoft .NET C#. A simple GUI interface was developed for the software to allow for ease of use (Figure A.4). The built-in timer component provided by the .NET framework is used in the implementation of the discrete event simulation technique discussed in Section A.1.3. At each "tick" interval, the simulation of the five vital physiological signs will take place. The default time interval is one second however, this can easily be adjusted by the user. For the generation of the vital physiological signs, a modified version of the Random Number Generator class library [174] was used.

Within the class library, five classes are implemented, one for each vital physiological sign. An object-oriented programming approach was taken, whereby all five vital sign classes inherit from the abstract class "VitalSign". Using this approach, a instance of each vital sign can be created and calling the *Generate* method enables the generation of a vital sign. The *Generate*

method requires the passing of two variables, the mean and standard deviation. The method then carries out the normal distribution calculations and returns a value, i.e. the vital physiological sign.

A.3 Evaluation of the Cloud4Health Platform

Prior sections has presented the design and implementation of the patient simulator. The sections which follow demonstrate the capabilities of using simulated patient data for the technical evaluation of the Cloud4Health platform. The methodology and presentation of results were originally presented in [175] and used here with permission from the publisher. An overview on the Cloud4Health project is first given.

A.3.1 Overview of Cloud4Health

The aim of Cloud4Health project is to develop, implement, validate and disseminate a novel, secure e-Health platform for capture, storage and consumption of data within a health care domain [1]. The key components of the Cloud4Health platform include:

- *Single Point of Contact (SPoC)* - An authorisation gateway for a health care domain to grant role-based access rights to sensitive medical data and services using a policy syntax.
- *Information Sharing Policy Syntax* - This is a rule-based language syntax inspired by firewall rules. It can be used to define access rights and to express a variety of patient consents.
- *Data Buckets* - The Data Buckets offer long-term persistence of medical data and support the Creation, Reading, Updating and Deletion (CRUD) of attribute values and associated meta data.

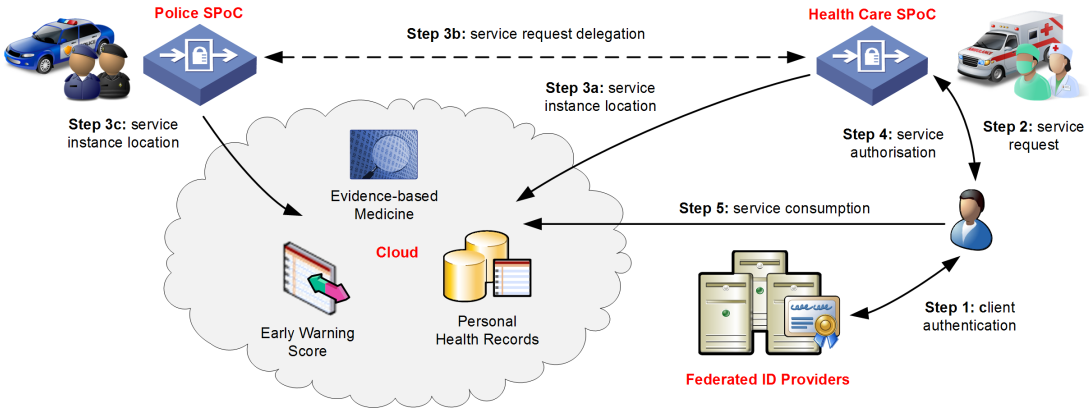

Figure A.5: Overview of the DACAR e-Health service platform

Figure A.5 gives an overview of the DACAR platform. Typically, a user consumes an e-Health service developed on the DACAR platform in five steps:

1. **Authentication**: The user logs on from federated identity providers using a user name and a password, or other unique personal information.

2. **Request for a service**: The user's client software forwards the security credential obtained in Step 1 to a responsible SPoC, together with a service request.

3. **Instantiate the service**: The SPoC checks the user's identity, resolves it into a role, and matches the service request to existing security policies. In the case that the service is provided by the local domain, the SPoC is able to tell whether the user is allowed to consume this service, and to locate the service endpoint within the Cloud. However, if the service is provided by a trustworthy foreign domain, the SPoC will route the service request to another SPoC. For example, when a clinician needs to make contact with a patient's relatives in an emergency, he or she sends a request for a police registry service to the local health care SPoC, which forwards the request to a remote police

SPoC.

4. **Authorisation**: If the service request is permitted by corresponding security policies, the SPoC that made the decision creates and signs a Service Ticket. This contains the user's pseudonym and role, a reference to the service endpoint, period of validity, and one-off session keys that enable the user's client software and a service instance to establish a secure session. Otherwise, a message is returned to tell the reason for rejection.

5. **Consume the service**: Finally, the user's client software initiates a secure session using the information provided in the Service Ticket and starts to consume the service. If the service requires CRUD operations over certain attributes, the service itself becomes a consumer of related Data Bucket services. In this case, the service needs to go through Steps 1 to 4 to obtain necessary Data Tickets from a SPoC using the service's own identity, or the service consumer's identity and role. In the latter circumstance the service is "impersonated", and shall use the Service Ticket received from its consumer as a complementary security credential.

Although this five step process in provides a secure environment for medical data, the limitation in this work is attempting to validate such claims. As described earlier, the DACAR implementation allows for the capture, storage and retrieval of data in a highly secured cloud environment but the solution of evaluating DACAR, and similar e-Health platforms, are still lacking. Thus, this chapter aims to assess whether meaningful evaluation results may be obtained from the Cloud4Health platform in conducting performance evaluation using the patient simulator prototype. The next section provides overview on related areas of work in regards to e-Health evaluation.

A.3.2 Performance Evaluation Metrics Design

As the name implies, Performance Evaluation focuses on assessing an e-Health implementation from a technical standpoint. The model aims to evaluate the scalability, functionality and reliability of a chosen e-Health implementation. Scalability relates to the volume of patient data interactions, i.e. uploading, downloading and processing of data, which the implementation can handle concurrently whilst functionality aims to assess how well the implementation works. Reliability aims to answer the question the integrity of an implementation in ensuring no data loss occurs during the evaluation process a common situation which occurs when a system is under high load and begins dropping data packets that it is unable to process.

The following is an outline of performance evaluation metrics which are defined in this work:

- CPU Utilization: relates to how much processing time is required for the upload and/or download of health care data along with general interaction with the e-Health implementation. The measurement of this metrics provides us with an overview on whether the current hardware infrastructure is up to a sufficient standard for the hosting of a chosen e-Health solution. Results obtained via this metric are dependent on the processor(s) of the hosting platform. CPU utilization is measured under the unit of percentage (%).

- Packet Loss: relates to the number of healthcare data samples that are lost or dropped during an evaluation. A low packet loss (preferably zero) is highly desirable, if not essential, for e-Health implementations since the key attribute in which all interactions revolve around is healthcare data.

- Upload / Download Time: the duration of time taken for a health-

care data to be uploaded or retrieved from an e-Health solution. Results obtained from this metric are dependent on the network interface cards (NIC) and protocols used by the e-Health implementation under evaluation. Measured in units of ms (milliseconds).

- Round-Trip Time (RTT): the duration of time taken for healthcare data to be uploaded to an e-Health platform, processed and then outputted to an e-Health service. As before, results from this metric are also dependent on NICs and the protocol used by the e-Health implementation under evaluation. The metric of RTT enables us to assess how well an e-Health implementation works for real-time scenarios such as uploading of healthcare data to a platform, processing of the data and outputting the data to an end clinical service, e.g. a patient monitoring system. A low RTT is obviously preferable in such scenarios. Measured in units of ms (milliseconds).

A.3.3 Evaluation Experiments

As noted earlier, evaluation is conducted on the Cloud4Health implementation whilst data and interaction is provided by the patient simulator prototype. The Cloud4Health project provides an API in order for the simulated data to interact with the platform but the underlying hardware architecture or source code is not considered during evaluation, i.e. a black-box evaluation. It's proposed that this helps provide a non-bias evaluation of the platform along with ensuring the defined evaluation metrics are not restricted to evaluating only one specific e-Health implementation for future work. For the scope of experimentations carried out in this chapter, the Cloud4Health platform is hosted within a sandbox virtual environment.

A.3.3.1 Methodology

Two primary sets of experiments in regards to performance evaluation of the Cloud4Health platform. Both experiments were conducted using simulated patient data generated on-the-fly at the start of each test. In each experiment, the only variation is the number of patients simulated. Experiment 1 aims to measure the performance evaluation metric of Upload Time whilst Experiment 2 measures the metrics of RTT and CPU utilization. Furthermore, packet loss is monitored in all experiments. The scenario for each experiment is outlined as follows:

- Experiment 1: Baseline Test - The first test is very simplistic in nature. In simulating 100 samples of a single patient's data, the upload time was monitored for intervals of 0.5, 1 and 3 seconds. In other words, 100 samples of a single patients data was uploaded with three different time delays in order to evaluate whether any performance impact was found on the Cloud4Health platform based on how "talkative" the client, i.e. Patient Simulator, is. The primary aim of this experiment is establish a baseline result for how well the Cloud4Health platform handles a single patient's data being uploaded.

- Experiment 2: RTT and CPU Utilization - The second experiment aimed to evaluate the RTT and CPU utilization of the Cloud4Health platform. Up to 100 patient's data was uploaded concurrently. Each patient simulates 100 samples of data. The aim of this experiment is to assess the Cloud4Health platform's latency under a more realistic scenario which involves the input, processing and output of patient data. Figure A.6 provides the workflow of the second experiment.

The methodology for gathering of defined metrics is as follows:

- CPU Utilization: A Microsoft Powershell script running directly on Cloud4Health server was used to obtain this metric. The script mon-

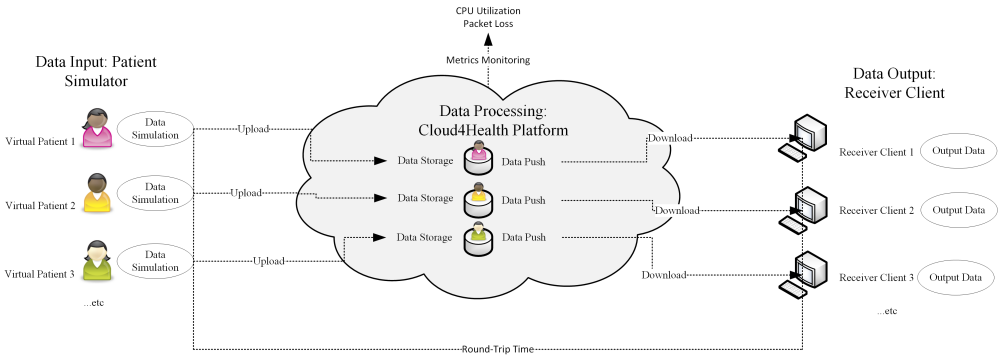

Figure A.6: Workflow of Experiment Two

itors the counter referred to as \processor(_total)% processor time [176]. This counter returns the overall CPU utilization of the server and the current value is logged to an output file every 1 second interval.

- Packet Loss: The total current samples stored in the Cloud4Health platform for each virtual patient was noted prior to the start of an experiment. The total samples simulated (100 per patient) is then subtracted from the current samples for each patient. A numeric value greater than 0 gives indication on the number of packets lost during the experiment.

- Upload Time: Upload time is gathered directly from the Patient Simulator application via the StopWatch class [177] provided by .NET C#. An instance of the StopWatch class is started upon uploading of data, and once the upload operation is complete, the StopWatch is stopped. The elapsed time produced by the StopWatch class results in the time taken to upload a single patient's data to the Cloud4Health platform.

- RTT: A push notification [178] service was implemented on top of the Cloud4Health platform. Via the implementation of a Receiver client (acting as an example clinical service), the RTT metric is calculated based on subtracting the time stamp a packet was received (by the Receiver service) against the time stamp of when a patient data

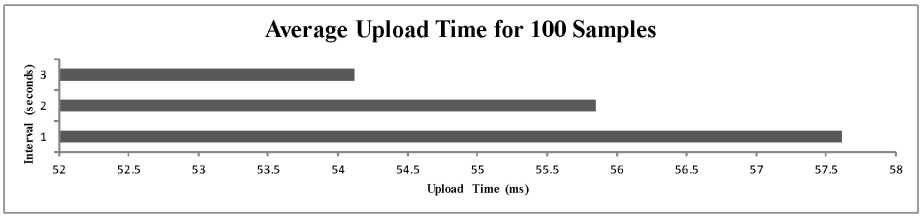

Figure A.7: Average Upload Time

sample was sent to the e-Health platform (via the Patient Simulator). Strong time synchronization was achieved via an Active Directory server acting as time synchronizer [179] between the Patient Simulator and Receiver client machines.

A.3.3.2 Results

Figure A.7 provides the average upload time for 100 samples of a single virtual patient's data uploaded at intervals of 0.5, 1 and 3 seconds. No packet loss occurred during the running of this experiment. As the interval time decreases, the time taken to upload a single virtual patient's data increases. Two conclusions can be made from this first experimental result: 1) talkativeness of a client, when upload a single patient's data, only affects the Cloud4Health platform in a very minor manner and 2) upload times for a single patient's data is exceptionally good with results of less than 58 ms when uploading data in intervals of 0.5 seconds.

As part of the second experiment, performance evaluation results of RTT latency and CPU utilization of the Cloud4Health platform are presented. No packet loss occurred in the instance of running this experiment. Figure A.8 shows the average RTT latency when simulating and uploading 20, 60 and 100 patient's data whilst Figure A.9 shows the CPU Utilization results. Due to wide variance, Table A.2 is also presented to give an overview of the minimum and maximum RTT latency values gathered during this experiment.

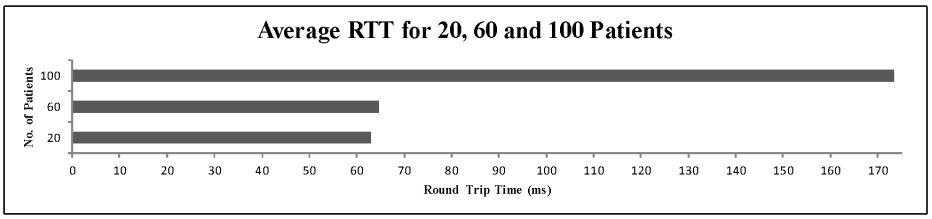

Figure A.8: Average Round-Trip Time

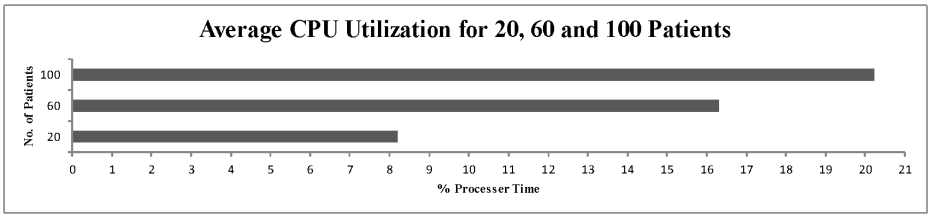

Figure A.9: Average CPU Utlisation

The results from the RTT graph show that in the case of simulating and uploading data for 20 virtual patients concurrently, the average RTT latency was very reasonable at 62.93 ms. In simulating 60 virtual patients data concurrently, the average RTT latency was found to be 64.80 ms which is only a very minor increase in comparison with 20 patients. On the other hand, simulating 100 patients produced a significantly higher latency of 173.56 ms on average. Furthermore, with a maximum latency of 6674.21 ms and minimum latency of 51.11 ms, there is far wider range of variance in RTT latency when simulating 100 patients. This wide range of variance is an indication that a bottleneck may be present on the Cloud4Health platform. CPU utilization was the primary suspect in this increased variance with RTT latency but Figure 4 shows this is not the case. Even in the scenario of simulating 100 patients, the average CPU utilization of the Cloud4Health platform was only slightly greater than 20%. Hence, it can be stated from this experiment that although RTT latency values grow as the number of patients simulated increase there is currently no direct evidence to show that this has any correlation with CPU utilization.

Overall, the key findings in conducting a performance evaluation of the

Table A.2: Minimum and Maximum Latency Values

Number of Patients	Min RTT (ms)	Max RTT (ms)
20	32.57	526.65
60	31.33	578.02
100	51.11	6674.21

Cloud4Health platform based on the criteria of functionality, reliability and scalability using the proposed metrics is as follows:

- Functionality: Both experiments have proven the functionality of this platform. The main goal of the Cloud4Health platform is to enable the storage of patient data with strong focus on security. Using simulated patient data, the experiments have proven that the implementation is capable of handling both the uploading and downloading of patient data whilst conducting authentication and authorization via security protocols.

- Reliability: In both experiments conducted, no packet loss occurred between the patient simulator and the Cloud4Health platform. Hence, it can be stated that the implementation is very reliable in ensuring that patient data is retained during any general interaction.

- Scalability: Scalability is perhaps the current primary limitation in the Cloud4Health implementation. Experimentation results were stopped after simulation of 100 patients data as it was found the Cloud4Health platform was unable to handle any higher volumes of patient data without unexpected errors. Though the Cloud4Health platform has been resourceful in the usage of CPU utilization, there is still a wide variance in the RTT latency values when processing 100 virtual patients data concurrently hence scope for improvement and optimization to ensure this platform is capable of handling higher volumes of patient data.

Appendix B

Probability Distribution

EXAMPLE of calculation of probability distribution's Probability Density Function (PDF) and Cumulative Distribution Function (CDF) is given in Section C.1. The normal distribution is used. Calculations are conducted in R ver. 3.01. Note that "[1]" indicates the output of the previous command given for each example in R.

B.1 PDF and CDF Calculation Example

The CDF of normal distribution is:

$$\frac{1}{2}[1 + erf(\frac{x-\mu}{\sqrt{2\sigma^2}})] \tag{B.1}$$

whilst the PDF is:

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{B.2}$$

PDF produces the density of a variable, which there is no quantifiable unit, whilst the CDF is between 0 and 1 thus multiplication of 100 produces a percentage. Consider parameters $\mu = 0$, $\sigma = 1$ and $x = 1$. Calculation of

CDF for $x = 1$, for $P(a < x < b)$, using intervals of $a = 0.99$ and $b = 1.01$, using R is as follows:

```
a = 0.5 * (1+erf((1.01-0)/sqrt(2*(1^2))))
[1] 0.8437524
b = 0.5 * (1+erf((0.99-0)/sqrt(2*(1^2))))
[1] 0.8389129
```

therefore:

```
a - b
[1] 0.004839414
```

where the error function, erf is defined as:

```
erf <- function(x) 2 * pnorm(x * sqrt(2)) - 1
```

The result for this calculation is $a - b = 0.004839414$. In other words, 0.48% probability of the variable 1 occurring in a normal distribution given the defined parameters. The calculation of PDF for the same parameters, $\mu = 0$, $\sigma = 1$ and $x = 1$, is as follows:

```
pdf = 1/(sqrt(2*pi)) * exp(-1^2/2)
[1] 0.2419707
```

which results in $pdf = 0.2419707$. To reiterate, the probability of variable X is 0.004839414 whilst its density is 0.2419707. Given that the interval range is between 0.99 to 1.01 the probability of X can be inferred from the PDF by multiplying it by the $Difference = 1.01 - 0.99$. Therefore:

$$0.2419707 * 0.02 = 0.004839414 \tag{B.3}$$

This shows that the PDF is the derivative of the CDF. Furthermore, the results may be easily verified with R's inbuilt library for normal distribution as follows. For CDF and PDF respectively:

```
cdf = pnorm(1.01, 0, 1) - pnorm(0.99, 0, 1)
[1] 0.004839414
pdf = dnorm(1, 0, 1) * 0.02
[1] 0.004839414
```

Both cdf and pdf will result in value of 0.004839414.

Appendix C

Anderson-Darling Test

EXAMPLE of calculating the Anderson-Darling (AD) test statistic is given in this section using R. Manual calculations, via R ver. 3.0.1, is compared to the automatically generated value given by Minitab 17 to provide proof that both methods are accurate in producing the AD test statistic. Note that "[1]" indicates the output of the previous command given for each example in R.

C.1 AD Test Statistic Calculations

The AD test statistic is calculated as follows:

$$A^2 = -n - S, \tag{C.1}$$

where variable S is derived as follows:

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln(F(Y_i)) + \ln\left(1 - F(Y_{n+1-i})\right)\right] \tag{C.2}$$

Suppose variable $Y = \{0, 0, 1, 2, -1, -1, 1, -2, 0\}$ comes from a normal distribution based on parameters $\mu = 0$ and $\sigma = 1.225$. The AD test statistic may be calculated by first solving $F(Y_i)$ and $1 - F(Y_{n+1-i})$. In both instances, F

is the Cumulative Distribution Function (CDF) of normal distribution give as:

$$\frac{1}{2}[1 + erf(\frac{x-\mu}{\sqrt{2\sigma^2}})] \tag{C.3}$$

The CDF of each value in Y is given in Table C.1. An example of calculation for the first row of results is presented below using R:

```
pnorm(-2,0,1.225) # F(Y_i)
[1] 0.05127099
1-pnorm(2,0,1.225) # 1 − F(Y_{n+1−i})
[1] 0.05127099
```

Next, the calculation of $\ln(F(Y_i)) + \ln(1 - F(Y_{n+1-i})$ can be achieved as shown in Table C.2. As before, example of the first row is given in R as follows:

```
log(0.05127099)+log(0.05127099) #ln(F(Y_i)) + ln(1 − F(Y_{n+1−i})
[1] -5.94126
```

Each of the results from the previous table can then be multiplied by $\frac{2i-1}{n}$ as the results show in Table C.3. The R calculation for the first row of result is:

```
((2*1-1)/9)* -5.94126 # (2i−1)/n [ln(F(Y_i)) + ln(1 − F(Y_{n+1−i})]
[1] -0.66014
```

The $\sum_{i=1}^{n}$ of each result in the previous table gives $S = -9.26094$. Thus, $A^2 = (-9) - (-9.26094) = 0.26094$. As shown in Figure C.1, the same AD test statistic result (rounding to 3 decimal places) is calculated by Minitab thus providing evidence that both techniques, manual calculation and the use of Minitab, for obtaining the AD test statistic is valid.

Table C.1: Result of $F(Y_i)$ and $1 - F(Y_{n+1-i})$

Y_i	$F(Y_i)$	$1 - F(Y_{n+1-i})$
-2	0.05127099	0.05127099
-1	0.2071567	0.2071567
-1	0.2071567	0.2071567
0	0.5	0.5
0	0.5	0.5
0	0.5	0.5
1	0.7928433	0.7928433
1	0.7928433	0.7928433
2	0.948729	0.948729

Table C.2: Results of $\ln(F(Y_i)) + \ln(1 - F(Y_{n+1-i})$

Y_i	$[\ln(F(Y_i)) + \ln(1 - F(Y_{n+1-i})]$
-2	-5.94126
-1	-3.14856
-1	-3.14856
0	-1.386294
0	-1.386294
0	-1.386294
1	-0.4642594
1	-0.4642594
2	-0.1052642

Table C.3: Results of $\frac{2i-1}{n}[\ln(F(Y_i)) + \ln(1 - F(Y_{n+1-i})]$

Y_i	$\frac{2i-1}{n}[\ln(F(Y_i)) + \ln(1 - F(Y_{n+1-i})]$
-2	-0.66014
-1	-1.04952
-1	-1.74920
0	-1.07823
0	-1.38629
0	-1.69436
1	-0.67060
1	-0.77377
2	-0.19883

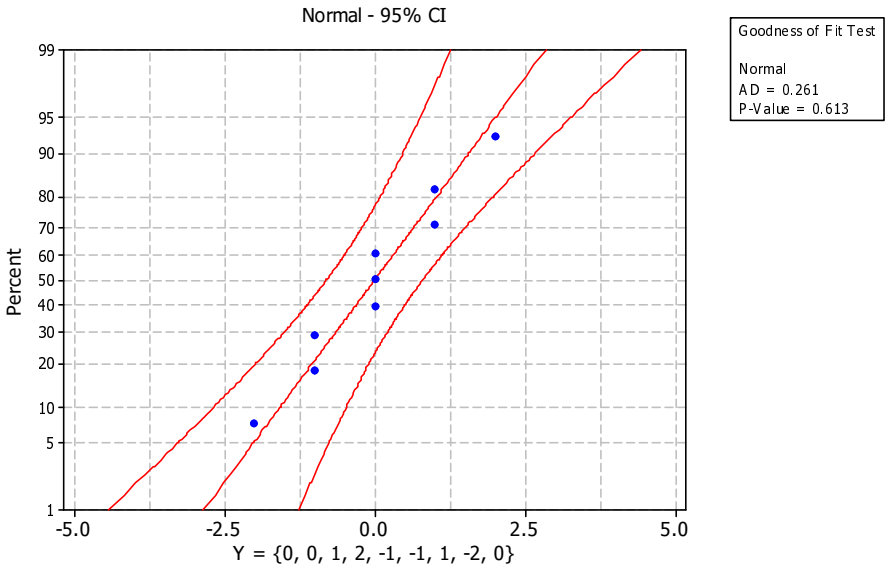

Figure C.1: AD Test Statistic Calculation by Minitab

Appendix D

Scripts

SECTION D.1 provides the R script used for modelling of each patient's one and two minute RR interval data based on the Fantasia Database. Section D.2 provides the script used to calculate parameters based on the Expectation Maximization (EM) algorithm implemented in the library "mixtools". Section D.3 provides the R script used for quartile based learning whist the decision making process is shown in Section D.4. Section D.5 and D.6 gives source code to the electronic version of the National Early Warning Score (NEWS) and Modified Early Warning Score (MEWS) algorithm for risk assessment of heart data respectively.

D.1 Individual Distribution Modelling Script

```
#f1o01 Normal Model
set.seed(100)
f1o01 <- rnorm(123,0.9692,0.04389)

#f1o02 Gamma Model
set.seed(100)
f1o02 <- rgamma(122,2296.39891,scale=0.00043)
```

```
#f1o03 Weibull* Model
set.seed(100)
f1o03 <- rweibull(124,25.17597,0.98128)

#f1o04 Weibull Model
set.seed(100)
f1o04 <- rweibull(96,34.64655,1.23427)

#f1o05 Weibull* Model
set.seed(100)
f1o05 <- rweibull(115,52.45041,1.04237)

#f1o06 Normal Model
set.seed(100)
f1o06 <- rnorm(99,1.19325,0.02648)

#f1o07 Logistic Model
set.seed(100)
f1o07 <- rlogis(121,0.9827,0.0167)

#f1o08 Gamma Model
set.seed(100)
f1o08 <- rgamma(153,636.92991,scale=0.00123)

#f1o09 Gamma Model
set.seed(100)
f1o09 <- rgamma(85,296.59721,scale=0.00473)

#f1o10 Normal Model
set.seed(100)
f1o10 <- rnorm(135,0.88009,0.0409)
```

```
#f1y01 Logistic Model
set.seed(100)
f1y01 <- rlogis(155,0.76866,0.03145)

#f1y02 Normal Model
set.seed(100)
f1y02 <- rnorm(130,0.90868,0.10745)

#f1y03 Weibull* Model
set.seed(100)
f1y03 <- rweibull(131,12.96976,0.94428)

#f1y04 Weibull Model
set.seed(100)
f1y04 <- rweibull(90,12.24812,1.3771)

#f1y05 Weibull Model
set.seed(100)
f1y05 <- rweibull(126,26.66953,0.96275)

#f1y06 Weibull Model
set.seed(100)
f1y06 <- rweibull(121,14.87207,1.02397)

#f1y07 Normal Model
set.seed(100)
f1y07 <- rnorm(107,1.11738,0.13236)

#f1y08 Weibull Model
set.seed(100)
f1y08 <- rweibull(128,14.00713,0.96614)
```

```
#f1y09 Weibull Model

set.seed(100)

f1y09 <- rweibull(144,16.20404,0.8534)


#f1y10 Gamma* Model

set.seed(100)

f1y10 <- rgamma(153,232.26905,scale=0.00335)
```

D.2 Mixture Modelling Script

```
modelMixture <- function(dataVector)

{

require(mixtools)

DataVector = na.omit(dataVector) #trim NA values

N = length(DataVector)


set.seed(100)

mixmdl = normalmixEM(DataVector, k=2)


probs = c(mixmdl$lambda[1])

dists = runif(N)

data = vector(length=N)


for(i in 1:N)

{

 if(dists[i]<probs[1])

{

data[i] = rnorm(1, mean=mixmdl$mu[1], sd=mixmdl$sigma[1])

}

else

{
```

```
data[i] = rnorm(1, mean=mixmdl$mu[2], sd=mixmdl$sigma[2])
}
}
plot(mixmdl,which=2)
return(data)
}
```

D.3 Quartiles Based Learning Script

```
Learning <- function(training)
{
trainingData = na.omit(training)
n = length(trainingData)
sortedValues = sort(60/trainingData)
q2 = median(sortedValues)

if ( n %% 2 == 0){ #even therefore we ommit median in calculations
index = n/2
q1 = median(head(sortedValues, index))
q3 = median(tail(sortedValues, index))
}
else # odd value therefore we include the median in calculations
{
index = ((n+1)/2)
q1 = median(head(sortedValues, index))
q3 = median(tail(sortedValues, index))
}

iqr = q3 - q1

lThresholdLowRisk = q1 - 1.5 * (iqr)
```

```
uThresholdLowRisk = q3 + 1.5 * (iqr)

lThresholdMedRisk = q1 - 2.25 * (iqr)
uThresholdMedRisk = q3 + 2.25 * (iqr)

lThresholdHighRisk = q1 - 3 * (iqr)
uThresholdHighRisk = q3 + 3 * (iqr)

results = list(Q1=q1, Q2=q2, Q3=q3, IQR=iqr,
    LThresholdLowRisk=lThresholdLowRisk,
    UThresholdLowRisk=uThresholdLowRisk,
    LThresholdMedRisk=lThresholdMedRisk,
    UThresholdMedRisk=uThresholdMedRisk,
    LThresholdHighRisk=lThresholdHighRisk,
    UThresholdHighRisk=uThresholdHighRisk)

return(results)
}
```

D.4 Quartile Approach for Heart Data Risk Assessment

```
#trainingResults is the vector which comes from the Learning() method
QuartileDetection <- function(modelDataFile, trainingResults)
{
model <- na.omit(modelDataFile)
n = length(model)
lowRiskScore <- 0
mediumRiskScore <- 0
highRiskScore <- 0
```

```
for (i in 1:n)
{
hr = 60/modelDataFile[i]
if (hr > trainingResults$UThresholdLowRisk && hr <=
    trainingResults$UThresholdMedRisk)
{
increment(lowRiskScore)
}
else if (hr > trainingResults$UThresholdMedRisk && hr <=
    trainingResults$UThresholdHighRisk)
{
increment(mediumRiskScore)
}
else if (hr > trainingResults$UThresholdHighRisk)
{
increment(highRiskScore)
}
else if (hr < trainingResults$LThresholdLowRisk && hr >=
    trainingResults$UThresholdMedRisk)
{
increment(lowRiskScore)
}
else if (hr < trainingResults$LThresholdMedRisk && hr >=
    trainingResults$LThresholdHighRisk)
{
   incement(mediumRiskScore)
}
else if (hr < trainingResults$LThresholdHighRisk)
{
increment(highRiskScore)
}
```

```
}

results =
    list(LowRiskScore=lowRiskScore,MediumRiskScore=mediumRiskScore,
HighRiskScore=highRiskScore)
return(results)
}
```

D.5 NEWS Algorithm for Heart Data Risk Assessment

```
NEWSDetection <- function(modelDataFile)
{
model = na.omit(modelDataFile)
n = length(model)

lowRiskScore <- 0
mediumRiskScore <- 0
highRiskScore <- 0

for (i in 1:n)
{
hr = 60 / model[i]

if (hr >= 41 && hr <= 50)
{
increment(lowRiskScore)
}
else if (hr <= 40)
{
```

```
increment(highRiskScore)
}
else if(hr >= 91 && hr <= 110)
{
increment(lowRiskScore)
}
else if(hr > 110 && hr <= 130)
{
increment(mediumRiskScore)
}
else if(hr > 130)
{
increment(highRiskScore)
}
}

results =
    list(LowRiskScore=lowRiskScore,MediumRiskScore=mediumRiskScore,
HighRiskScore=highRiskScore)
return(results)
}
```

D.6 MEWS Algorithm for Heart Data Risk Assessment

```
MEWSDetection <- function(modelDataFile)
{
model = na.omit(modelDataFile)
n = length(model)
```

```
lowRiskScore <- 0

mediumRiskScore <- 0

highRiskScore <- 0


for (i in 1:n)

{

hr = 60 / model[i]


if (hr >= 41 && hr <= 50)

{

increment(lowRiskScore)

}

else if (hr <= 40)

{

increment(highRiskScore)

}

else if(hr >= 101 && hr <= 110)

{

increment(lowRiskScore)

}

else if(hr > 110 && hr <= 129)

{

increment(mediumRiskScore)

}

else if(hr > 129)

{

increment(highRiskScore)

}

}


results =

    list(LowRiskScore=lowRiskScore,MediumRiskScore=mediumRiskScore,
```

```
HighRiskScore=highRiskScore)
return(results)
}
```

D.7 Hybrid Approach Script

```
HybridDetection <- function(modelDataFile, trainingResults)
{
model <- na.omit(modelDataFile)
n = length(model)
lowRiskScore <- 0
mediumRiskScore <- 0
highRiskScore <- 0

for (i in 1:n)
{
hr = 60/modelDataFile[i]
if (hr > trainingResults$UThresholdLowRisk && hr <=
    trainingResults$UThresholdMedRisk && hybridNEWS(hr) == TRUE)
{
increment(lowRiskScore)
}
else if (hr > trainingResults$UThresholdMedRisk && hr <=
    trainingResults$UThresholdHighRisk && hybridNEWS(hr) == TRUE)
{
increment(mediumRiskScore)
}
else if (hr > trainingResults$UThresholdHighRisk && hybridNEWS(hr)
    == TRUE)
{
increment(highRiskScore)
```

```
}
else if (hr < trainingResults$LThresholdLowRisk && hr >=
    trainingResults$UThresholdMedRisk && hybridNEWS(hr) == TRUE)
{
increment(lowRiskScore)
}
else if (hr < trainingResults$LThresholdMedRisk && hr >=
    trainingResults$LThresholdHighRisk && hybridNEWS(hr) == TRUE)
{
increment(mediumRiskScore)
}
else if (hr < trainingResults$LThresholdHighRisk && hybridNEWS(hr)
    == TRUE)
{
increment(highRiskScore)
}
}

results =
    list(LowRiskScore=lowRiskScore,MediumRiskScore=mediumRiskScore,
HighRiskScore=highRiskScore)
return(results)
}

#NEWS logic for heart rate risk assessment for hybrid approach to
    query. Simply true or false is returned rather than a risk score
hybridNEWS <- function(hr)
{
if (hr >= 41 && hr <= 50)
{
return(TRUE)
}
```

```
else if (hr <= 40)
{
return(TRUE)
}
else if(hr >= 91 && hr <= 110)
{
return(TRUE)
}
else if(hr > 110 && hr <= 130)
{
return(TRUE)
}
else if(hr > 130)
{
return(TRUE)
}
else
{
return(FALSE)
}
}
```

D.8 Increment Function

Simple increment script required for all scripts which call the *increment*() function:

```
increment <- function(x)
{
eval.parent(substitute(x <- x + 1))
}
```

Appendix E

EM Algorithm Parameter Estimates

TABLE E.1 provides the parameter estimates of each elderly patient using the Expectation Maximization (EM) algorithm. Table E.2 provides the parameter estimate of each young patient using the EM algorithm. The script used to conduct EM algorithm estimates is provided in Appendix D. Rounding of 5 decimal places has been applied to the reported values.

Table E.1: EM Parameter Estimate (Elderly Patients)

Dataset	**Mean μ**	**SD σ**	**Weight %**
f1o01	0.906303	0.015141	0.182245
	0.983221	0.034748	0.817755
f1o02	0.97246	0.017809	0.898815
	1.010039	0.003689	0.101185
f1o03	0.936703	0.060282	0.388592
	0.976179	0.024616	0.611408
f1o04	1.199323	0.036064	0.707871
	1.253985	0.017497	0.292129
f1o05	1.013225	0.027479	0.456655
	1.045795	0.009309	0.543345
f1o06	1.138268	0.014184	0.043718
	1.195766	0.023921	0.956282
f1o07	0.967736	0.044999	0.287414
	0.986617	0.021235	0.712586
f1o08	0.770286	0.023633	0.799428
	0.822888	0.019012	0.200572
f1o09	1.402691	0.088082	0.615614
	1.39973	0.06855	0.384386
f1o10	0.808166	0.009368	0.086304
	0.886882	0.035694	0.913696

Table E.2: EM Parameter Estimate (Young Patients)

Dataset	**Mean μ**	**SD σ**	**Weight %**
f1y01	0.710698	0.004272	0.027698
	0.773136	0.058045	0.972302
f1y02	0.90013	0.101975	0.954434
	1.087708	0.006257	0.045566
f1y03	0.752207	0.013877	0.120294
	0.928089	0.066144	0.879706
f1y04	1.127704	0.042317	0.183187
	1.364759	0.092141	0.816813
f1y05	0.860498	0.039222	0.101778
	0.952684	0.033098	0.898222
f1y06	0.963736	0.077028	0.744367
	1.060594	0.023428	0.255633
f1y07	0.929623	0.049472	0.166833
	1.15498	0.108933	0.833167
f1y08	0.842623	0.061277	0.330639
	0.973955	0.047506	0.669361
f1y09	0.80664	0.070348	0.530832
	0.849503	0.031852	0.469168
f1y10	0.760249	0.02087	0.316117
	0.785919	0.058684	0.683883

Appendix F

Comparison between NEWS and MEWS Algorithm

THIS section provides evidence on the minor, non significant, variation which exists between the National Early Warning Score (NEWS) and Modified Early Warning Score (MEWS) algorithm when assessing heart rate only. Furthermore, it provides additional evidence to support the argument that there is sensitivity in the Early Warning Score (EWS) algorithms due to risk scores being predefined in a static manner based on vital sign parameter values observed. A subset of the first five elderly patients and young patients from the Fantasia dataset is used in this experiment. Each real-life recording is converted from RR Sample to RR Intervals to Beats Per Minute. Recordings are played back to the NEWS and MEWS algorithm (see Appendix D.5 and Appendix D.6 for implementation) and the total number of false-positive are logged.

F.1 NEWS and MEWS Comparison

Table F.1 gives the scoring logic for heart rate in both the MEWS and NEWS algorithm whilst Table F.2 provides the results of this experiment.

Table F.1: MEWS and NEWS Algorithm Comparison

Parameter	3	2	1	0	1	2	3
MEWS Heart Rate	≤40		41-50	51-100	101-110	111-129	≥130
NEWS Heart Rate	≤40		41-50	51-90	91-110	111-130	≥131

Table F.2: MEWS and NEWS Heart Score Results

Dataset	Algorithm	Scores of 1	Scores of 2	Scores of 3	Accuracy Ratio
f1o01	MEWS	0	0	0	1.00
	NEWS	0	0	0	1.00
f1o02	MEWS	9	0	0	0.99
	NEWS	9	0	0	0.99
f1o03	MEWS	13	0	0	0.99
	NEWS	13	0	0	0.99
f1o04	MEWS	2253	0	1	0.64
	NEWS	2253	0	1	0.64
f1o05	MEWS	9	0	0	0.99
	NEWS	9	0	0	0.99
f1y01	MEWS	27	0	0	0.99
	NEWS	150	0	0	0.98
f1y02	MEWS	76	0	0	0.99
	NEWS	94	0	0	0.98
f1y03	MEWS	1	0	0	0.99
	NEWS	1	0	0	0.99
f1y04	MEWS	4253	0	256	0.18
	NEWS	4253	0	256	0.18
f1y05	MEWS	323	0	0	0.95
	NEWS	323	0	0	0.95

The most interesting results can be seen in Patient's f1o04 and f1y04 where the accuracy ratio is approximately 64% and 18 % respectively. Such a result indicates significant false positive ratio in the EWS algorithm when analysing these two patient's heart data. It can be seen that any value less than 50 Beats Per Minute (BPM) entails a risk score of 1. Although the nature of the triggers are mainly low risk scores, the high sensitivity in the overall results demonstrates that there is potential to fine tune the EWS algorithm for better accuracy of results. Comparatively, the results also show that risk assessment of heart rate in both the NEWS and MEWS algorithm are very similar. The only deviation exists in two of the tests ran which are found in Patient's f1y01 and f1y02. The NEWS algorithm triggered slightly higher number of false-positives due to the logic of defining normality as heart rate within the range of $51 - 90$ whilst the MEWS algorithm caters to a higher range of $51 - 100$.