



IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

---

## A real-time independent and inexpensive PPG signal quality classification tool for vital sign monitoring

---

*Author:*

Omar Muttawa

*Supervisor:*  
Dr. Jesus Rodriguez Manzano

*Co-Supervisor:*  
Dr. Pantelis Georgiou

A thesis submitted for the degree of  
*MEng Electrical and Electronic Engineering*

November 18, 2021

### **Plagiarism Statement**

- I affirm that I have submitted, or will submit, electronic copies of my final year project report to both Blackboard and the EEE coursework submission system.
- I affirm that the two copies of the report are identical.
- I affirm that I have provided explicit references for all material in my Final Report which is not authored by me and represented as my own work.

## **Abstract**

In modern day healthcare photoplethysmogram signals (PPG) are constantly proving to be a utile method for non-intrusively, continuously and inexpensively monitoring a patients' vital signs. PPG signals are sensitive to several artefacts introduced as a result of motion, ambient light changes and incorrect probe placement which will corrupt conclusions formed from their analysis. There is hence a need for a real time, universal and 'light weight' algorithm to detect all such artefacts, the development of which is what is investigated in this report. Several scores proposed in the literature were evaluated on their ability to classify segments of a PPG signal in real time into three different classes of quality. Skewness and a matching of multiple systolic wave detection algorithms score were found to allow for a 'light weight' linear classification of segments of PPG signals from datasets acquired from various PPG devices. To find optimal thresholds, several linear machine learning classifiers were investigated. The final algorithm was then integrated into a graphic user interface capable of evaluating signal quality from the AFE4900 PPG sensor in real time as well any imported signal and was developed to be used in an ongoing study by the Centre for Bio-inspired Technology.

### **Acknowledgements**

I'd first like to thank Dr. Jesus Rodriguez Manzano and Dr. Pantelis Georgiou for providing me with the opportunity to conduct this project. The project allowed me to apply much of what I had learnt as part of my degree to the rewarding and impactful field of bio-engineering and reaffirmed my decision to pursue a career in engineering four years ago.

I'd then like to thank Stefan Karolick for his invaluable guidance and support throughout the entirety of the project. Stefan always made time to answer any questions I had and helped shape my research whenever I was unsure of the next step.

Finally I'd like to thank my friends and family for all their support during my degree with special thanks to Konstantinos Barmpas who has always provided me with a source of inspiration over the last four years.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
<b>2</b>	<b>Background</b>	<b>7</b>
2.1	Importance of continuous physiological signal measurement . . . . .	7
2.2	Introduction to photoplethysmography . . . . .	7
2.3	Ongoing research at the Centre for Bio-inspired technology . . . . .	9
2.3.1	Measuring hematocrit using photoplethysmography . . . . .	9
2.3.2	Measuring blood pressure using photoplethysmography . . . . .	9
2.3.3	AFE4900EVM and sensor prototypes . . . . .	10
2.4	The difficulties of photoplethysmography acquisition and the need for automated quality classification . . . . .	11
2.5	Existing 'lightweight' approaches to PPG signal quality evaluation . . . . .	12
2.5.1	SQIs using electrocardiogram signals as reference . . . . .	12
2.5.2	PPG independent SQIs using a dynamic morphological template as reference	13
2.5.3	SQIs for independent real-time application with no calibration . . . . .	14
2.6	Background of proposed approach . . . . .	15
2.6.1	The pre-processing of photoplethysmogram signals . . . . .	15
2.6.2	Annotation method . . . . .	17
2.6.3	Features Extracted . . . . .	18
2.6.4	Man-Whitney U and Wilcoxon signed-rank tests . . . . .	19
2.6.5	Extra random forest classifier feature importance test . . . . .	20
2.6.6	The use of machine learning methods in linear classification . . . . .	21
2.6.7	Precision, Recall and the F1 Score . . . . .	21
2.6.8	K-fold Cross-validation . . . . .	21
<b>3</b>	<b>Requirements Capture</b>	<b>25</b>
3.1	Signal Quality Algorithm: . . . . .	25
3.1.1	Independence . . . . .	25
3.1.2	Universality . . . . .	25
3.1.3	Real-time application . . . . .	25
3.1.4	Utility . . . . .	25
3.1.5	Computational requirement . . . . .	26
3.2	Graphic User Interface: . . . . .	26
3.2.1	Compatibility . . . . .	26
3.2.2	Ease of Use . . . . .	26
3.2.3	Utility . . . . .	26
3.2.4	Export Capabilities . . . . .	26
<b>4</b>	<b>Analysis and Design</b>	<b>27</b>
4.1	Data Acquisition: . . . . .	27
4.1.1	The BIDMC Dataset . . . . .	27
4.1.2	Data collected at the HTD, Vietnam . . . . .	28
4.1.3	Data collected using the AFE4900EVM module . . . . .	28
4.2	Data Processing: . . . . .	28
4.3	Data annotation: . . . . .	30
4.4	SQI Analysis: . . . . .	31
4.4.1	Train, Validation and Test Split . . . . .	31

4.4.2	SQIs selected . . . . .	31
4.4.3	MSQ score implementation . . . . .	32
4.4.4	Computation time . . . . .	32
4.4.5	Mann-Whitney and Wilcoxon signed-rank tests . . . . .	33
4.4.6	Feature importance's using a Random Forest Classifier . . . . .	34
4.4.7	SQI selection . . . . .	35
4.5	Finding Linear Classification Thresholds . . . . .	35
4.5.1	Plot of training data in feature space . . . . .	36
4.5.2	Support Vector Machines . . . . .	36
4.5.3	Linear Classifiers with different loss functions . . . . .	37
4.5.4	Shallow Decision Tree . . . . .	39
4.5.5	Selected linear threshold . . . . .	39
4.5.6	Two-class classifier . . . . .	40
<b>5</b>	<b>Implementation</b>	<b>44</b>
5.1	Algorithm Implementation into Graphic User Interface . . . . .	44
5.2	Tools for Further Algorithm Development . . . . .	45
<b>6</b>	<b>Testing and Results</b>	<b>48</b>
6.1	Testing Accuracy of Implemented Algorithms . . . . .	48
6.1.1	Three-class classifier . . . . .	48
6.1.2	Two-class classifier . . . . .	48
6.2	Feature Space Distribution analysis . . . . .	49
6.3	Testing Computation Time of Implemented Algorithms . . . . .	50
6.4	Computational Complexity of Signal Quality Indices reviewed . . . . .	51
<b>7</b>	<b>Evaluation</b>	<b>52</b>
7.1	Context . . . . .	52
7.2	Meeting the set requirements . . . . .	52
7.2.1	Meeting the algorithm requirements . . . . .	52
7.2.2	Meeting the Graphic User Interface requirements . . . . .	53
7.3	Comparison of Findings to the Literature . . . . .	53
7.4	Effect of MSQ Score . . . . .	54
7.5	Optimal length Omission . . . . .	54
7.6	Precision and Accuracy of Speed Tests . . . . .	54
7.7	Draw Backs of Universality . . . . .	55
7.8	Effect of Annotation . . . . .	55
7.9	Dataset Size and Distribution of the dataset . . . . .	55
<b>8</b>	<b>Conclusions and Further Work</b>	<b>56</b>
<b>9</b>	<b>User Guide</b>	<b>58</b>
	<b>Bibliography</b>	<b>65</b>
	<b>A Figures</b>	<b>66</b>

# List of Figures

2.1	Diagram illustrating the basics of photoplethysmography . . . . .	8
2.2	Diagram illustrating the possible locations where PPG signals can be acquired [17]	8
2.3	Simplified Block Diagram of AFE4900 [35] . . . . .	10
2.4	Diagram illustrating the OPRA algorithm [48] . . . . .	13
2.5	Diagram illustrating Sun et al's proposed algorithm [50] . . . . .	14
2.6	Ideal frequency response for a butterworth filter [62] . . . . .	16
2.7	Three categories defined for PPG quality [56] . . . . .	17
2.8	Autocorrelation SQI illustration . . . . .	20
2.9	10-fold cross-validation illustration [84] . . . . .	22
2.10	Diagram illustrating support vector machine hyper-plane formulation [85] . . . . .	22
4.1	Examples of class A, B and C samples from the BIDMC dataset . . . . .	27
4.2	Examples of class A, B and C samples from the Vietnam HTD dataset . . . . .	28
4.3	Examples of class A, B and C samples from the AFE dataset . . . . .	29
4.4	Diagram of pre-processing pipeline . . . . .	29
4.5	Two dimensional Grid Search for Optimal MSQ parameter selection . . . . .	32
4.6	Computation time of all computed scores . . . . .	33
4.7	AB vs C Man-Whitney and Wilcoxon signed-rank test $p$ values for all SQIs . . . . .	34
4.8	A vs BC Man-Whitney and Wilcoxon signed-rank test $p$ values for all SQIs . . . . .	34
4.9	Extra trees mean relative importance of SQIs with standard deviations . . . . .	35
4.10	Feature space plot of training data . . . . .	36
4.11	SVM regularisation parameter investigation . . . . .	37
4.12	Linear SVM thresholds on training set . . . . .	37
4.13	SVM kernel investigation . . . . .	38
4.14	Linear classifier loss function investigation . . . . .	38
4.15	Fitted decision tree classification thresholds . . . . .	39
4.16	10-fold cross-validation F1 scores of the classifiers with standard deviation for three-class classifier . . . . .	40
4.17	10-fold cross-validation F1 scores of the classifiers with standard deviation for two-class classifier . . . . .	41
4.18	Squared loss two-class classifier segmentation . . . . .	41
4.19	Diagram illustrating three-class classifier algorithm . . . . .	42
4.20	Diagram illustrating two-class classifier algorithm . . . . .	43
5.1	Diagram of thread communication . . . . .	44
5.2	Annotation of signal acquired and analysed in real-time . . . . .	45
5.3	Annotation of signal imported and analysed . . . . .	46
5.4	Flow of GUI created . . . . .	46
5.5	GUI for dataset annotation . . . . .	47
6.1	Confusion matrix of test set for three-class classifier . . . . .	49
6.2	Confusion matrix of test set for two-class classifier . . . . .	49
6.3	MSQ class A example . . . . .	50
6.4	Computation time of algorithms . . . . .	50
A.1	Examples of class A samples from each subset raw and filtered . . . . .	66
A.2	Regularisation Parameter optimisation for varying three class classifiers . . . . .	67
A.3	Regularisation Parameter optimisation for varying two . . . . .	68

# List of Tables

2.1	Factors altering PPG response [38] . . . . .	11
4.1	Class distribution of acquired data . . . . .	30
4.2	SQI number table . . . . .	31
4.3	full table of results for three-class linear classification investigation . . . . .	40
4.4	Full table of results for two-class linear classification investigation . . . . .	41
6.1	Testing classification report of three-class classifier . . . . .	48
6.2	Testing Classification report of two-class classifier . . . . .	49

# Chapter 1

## Introduction

Following the events of the global COVID-19 pandemic the strain of infectious diseases on even the most developed and advanced healthcare systems have been brought to the forefront of general conversation. Although the scale of the COVID-19 pandemic is one that is unprecedented in modern medicine, countries regularly experience such strains on their healthcare systems in epidemic and endemic outbreaks of various infectious diseases.

The need to monitor the vital signs of the thousands of patients administered to hospitals during infectious disease outbreaks accounts for a large portion of this strain as current clinical approaches require the use of expensive, scarce and resource intensive equipment. For diseases such as Dengue Fever and Sepsis which are estimated to collectively impact the lives of 130 million people worldwide every year [1], [2], a inexpensive method to monitor patients vitals' continuously is urgently needed to facilitate the appropriate use of limited resources, improve case management and save lives.

Photoplethysmogram signals (PPG signals) are increasingly used in hospitals to continuously monitor the vital signs of patients. Their commercial adoption in fitness and health tracking devices which has reduced the price of the technology and ongoing research into the different vital signs that are capable of being measured using the technology presents healthcare professionals with a viable solution however the acquisition of these signals remains sensitive to several artefacts introduced as a result of motion, ambient light changes and incorrect probe placement. These artefacts limit the development and application of any methods which estimate vital signs from PPG signals in real time and therefore limit their clinical application.

This report outlines the development and implementation of a real-time independent and inexpensive PPG signal quality classification tool by first reviewing 26 existing signal quality indices proposed in the literature for their ability to independently segment PPG signal segments with different defined classes of Signal Quality. A inexpensive linear classification algorithm is then developed using 2 signal quality indices for real time detection, the thresholds of which were found by investigating several linear machine learning models. This algorithm was then implemented into a graphic user interface capable of recording and classifying signal segments in real-time as well as importing existing signals for classification. The outcomes of the project were subsequently evaluated with suggestions for further development identified.

# Chapter 2

## Background

### 2.1 Importance of continuous physiological signal measurement

When a patient is admitted to hospital their vital signs are often measured to assess their clinical situation. For critical patients these measurements are often repeated, as time passes, to identify if a patient is clinically deteriorating or improving. The measurements are often of a patient's heart rate (HR), blood pressure (BP), respiratory rate (RR) and blood oxygen saturation (SPO<sub>2</sub>) [3], the acquisition of which currently requires the use of expensive clinical monitoring equipment. As a direct consequence of this and the time required for a trained caregiver to acquire said measurements, the frequency of these measurements can often be sparse for a given patient.

For some conditions the continuous measurement of vital signs is of paramount importance when monitoring a patient's state, which may deteriorate very quickly as well as when attempting to identify key variations in the values of these vital signs which could provide an in-depth indication of a patient's current and future condition [3]–[7]. Dengue Fever, Sepsis as well as the COVID-19 pandemic are all examples of diseases that can severely strain a hospital's resources as a result of the absence of understanding the possible future severity of patients and an abundance of research is currently being conducted attempting to identify key indicators of these diseases' progression [8]–[10] such that to allow for the appropriate use of limited resources, improve case management and save lives.

To facilitate this research as well as to allow for a realistic large-scale application of its findings the factors that are hypothesised to be key to disease progression must be able to be acquired using inexpensive and simple methods [8]. This poses clinicians and researchers with a difficult challenge, however, ongoing research into the application of photoplethysmography to continuously measure several vital signs may provide a possible solution.

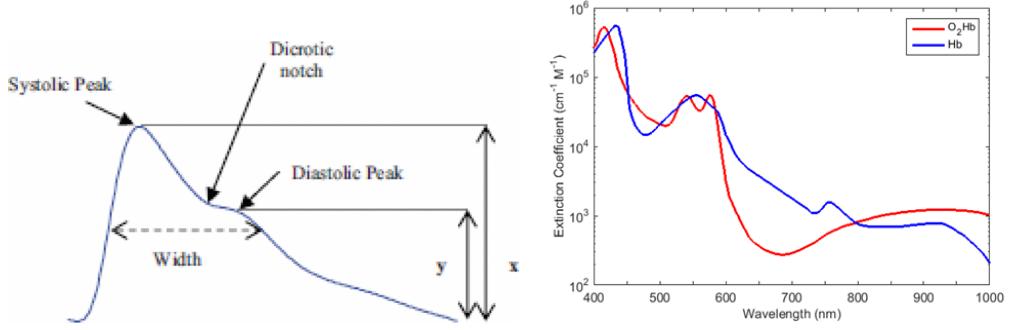
### 2.2 Introduction to photoplethysmography

Photoplethysmography (PPG) is a simple and low-cost optical non-invasive technique for detecting blood volume changes in a micro-vascular bed of tissue. It comprises of a pulsatile physiological waveform attributed to changes in blood volume with each heart beat, which is modulated by a slowly varying baseline attributed to respiration, sympathetic nervous system activity and thermoregulation [11]. Figure 2.1a shows the shape of a clear high quality PPG waveform where the systolic and diastolic peaks can be clearly identified as well as the corresponding discrotic notch.

A basic PPG sensor requires only a few opto-electronic components: a light emitting diode (LED) to illuminate the tissue with light of a specific wavelength and a photo-detector to measure the small variations in light intensity. The key factors that can affect the amount of light received by the photo-detector are blood volume, blood vessel wall movement and the orientation of red blood cells. [11]

The wavelength of the light transmitted is also very important to take into consideration for interactions between light and tissue for three main reasons:

- **The optimal water window:** Water, the main constituent of tissue, absorbs ultraviolet and longer infrared wavelengths very strongly. On the other hand, shorter wavelengths are



(a) Pulsatile component of a PPG signal [12] (b) Absorption spectra of  $\text{HbO}_2$  and  $\text{Hb}$  [13]

Figure 2.1: Diagram illustrating the basics of photoplethysmography

strongly absorbed by melanin. There is a window between these two boundaries that allows visible red and near infrared light to pass more easily, thereby facilitating the measurement of blood flow at these wavelengths [14].

- **Isobestic wavelength:** There are significant differences between the absorption spectrum of oxyhaemoglobin( $\text{HbO}_2$ ) and reduced haemoglobin ( $\text{Hb}$ ) shown in 2.1b. This allows for the non-invasive measurement of Blood oxygen content (SPO<sub>2</sub>) by measuring the ratio of light absorbed by tissue for a wavelength sensitive to  $\text{HbO}_2$  absorption and of the light absorbed by tissue for a wavelength sensitive to  $\text{Hb}$  absorption [15].
- **Tissue penetration depth:** The depth to which light penetrates the tissue for a given intensity of light depends upon the light's wavelength. [16]

The most common application of this theory is the use of an infrared light to measure volumetric variations of blood circulation and hence calculate ones heart-rate however in a clinical setting PPG sensors are often utilised which use both red and infrared light to non-invasively measure a patients blood oxygen content (SPO<sub>2</sub>) as explained above [15], [17].

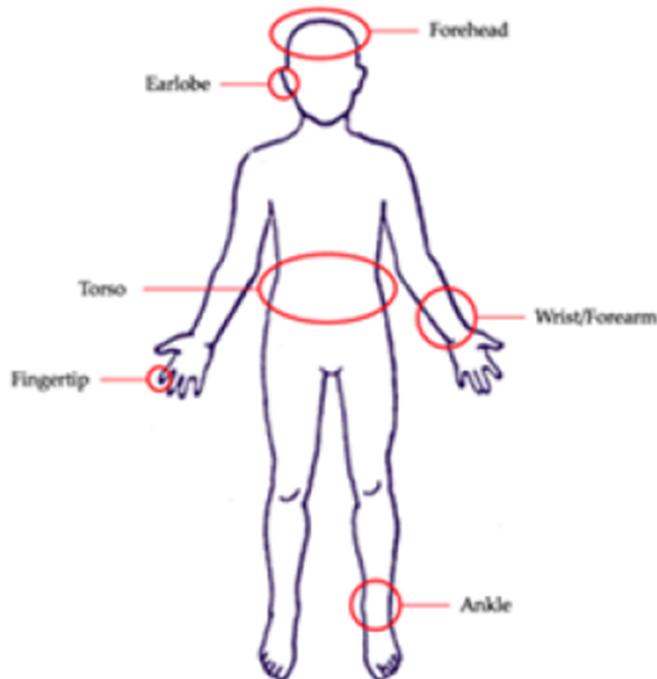


Figure 2.2: Diagram illustrating the possible locations where PPG signals can be acquired [17]

PPG sensors can only be placed on certain body locations as shown in Figure 2.2 which have all been proven utile for PPG acquisition, however, these locations will produce signals with varying

quality and each produce specific advantages and disadvantages for both their acquisition and their subsequent analysis [17], [18]. There are also two different distinct forms that a PPG sensor may have: transmission mode and reflectance mode. Each mode also posses its advantages and disadvantages in their acquisition and their subsequent analysis.

In transmission mode the detector and the light source are separated by the incident tissue. PPG signals from the fingertip and earlobe are often acquired using a transmission mode sensor as ambient light can be reduced by using a probe that encapsulates the fingertip or earlobe, however, too much pressure can slow down the peripheral blood volume, which can result in the reduction of the venous oscillations that produce a PPG signal [17].

In reflectance mode the photo-detector is positioned along the light source on the same side of the tissue to measure the reflected light. A reflectance mode sensor is required for the acquisition of PPG signals from ones wrists, forearms, ankles, forehead or torso [17]. For all such sensors the amount of pressure is once more paramount to acquisition such as to limit ambient light but also avoid a reduction in venous oscillation [19].

Although there are wavelengths, locations and acquisition modes that are much more common then others, sensors situated at different locations, using different wavelengths and acquisition modes continue to be created when attempting to measure novel clinical parameters or when attempting to optimise quality or robustness [20]–[26].

## 2.3 Ongoing research at the Centre for Bio-inspired technology

The Centre for Bio-inspired technology is conducting ongoing research in the use of PPG signals to continuously measure other previously non-feasible vital signs using PPG signals. The goal of this research and investigation is to attempt to monitor patients suffering from Dengue Fever and potentially predict their severity.

Dengue Fever (DF) is a flu-like illness spread by the bite of an infected female Eddie mosquito [27]. It is widespread among many of the tropical and subtropical regions of the world [28]. A patient with DF will often experience a sudden temperature increase accompanied by a headache and various other symptoms such as vomiting, abdominal pain and moderate thrombocyto-penia (low blood platelet count). If a DF patient experiences increasing vascular permeability (plasma leakage from blood vessels) they might progress to become a severe dengue patient and are at risk of experiencing Dengue Shock Syndrome (DSS) (weak rapid pulse, narrow pulse pressure, cold, clammy skin and restlessness) which can be fatal. [27]

The background of the approaches used to estimate these vital signs are outlined below.

### 2.3.1 Measuring hematocrit using photoplethysmography

Hematocrit is the percentage of red blood cells within the total volume of ones blood. It is clear in the literature that hematocrit values rising have strong relation to the risk of a patient entering DSS [8]. The realisation of which seems to suggest that one simply needs to continuously monitor a patients hematocrit to predict their risk of entering dengue shock. In reality however taking a patients' hematocrit requires a patients' blood to be sampled and then placed in a centrifuge to separate the red blood cells from the other components of blood. Continuously measuring hematocrit would hence require regular samples of a patients blood to be taken and centrifuged before being analysed. This is unrealistic during an epidemic where thousands of patients are admitted into hospitals each day.

The centre for Bio-inspired technology is currently investigating the use of PPG signals to estimate hematocrit optically in a similar method to that utilised to estimate oxygen saturation in blood. This research intends to extend upon an investigation conducted by Schmitt et Al [24] and validated by Ekuni et al [25] suggesting a linear relationship between the difference in log intensity of light received from PPG signals acquired at two near infra-red wavelengths equal to 800nm and 1300nm.

### 2.3.2 Measuring blood pressure using photoplethysmography

Detecting changes in blood pressure was also found in the literature to provide an indication of plasma leakage [29] and hence could be a key factor in diagnosing dengue fever. Currently

the conventional continuous measurement of blood pressure requires a patient to have an invasive arterial line put in to their arm attached to an expensive sensor that calculates BP. This is not only immensely discomforting but also very resource intensive, making the continuous measurement of blood pressure unrealistic.

A lot of research has been recently conducted to measure continuous non-invasive blood pressure (cNIBP) and success has been achieved for various estimation techniques using PPG signals independently, by extracting features from the PPG signal and using Machine learning methods [30], as well as using a reference electrocardiogram (ECG) signal to estimate CNIBP [23], [31] by measuring Pulse wave velocity found to be related to blood pressure by Moens, Korteweg (MK) [32] and Hughes [33].

A patent filed by Samsung in 2016 [34] suggests that blood pressure can be measured using a multiple wavelength PPG approach. Inventors at Samsung claim in the patent that the phase difference between the PPG signals received can be used to calculate Blood Pressure. The centre for Bio-inspired technology is investigating and reviewing these potential methods of continuous blood pressure estimation with the aim of being able to continuously track fluctuations in the vital signs value for monitored dengue patients.

### 2.3.3 AFE4900EVM and sensor prototypes

To conduct a review and investigation of the applicability of the research outlined above the AFE4900 device has been selected as the method of PPG signal acquisition. The AFE4900 is a ultra-low-power integrated analogue front end (AFE) analogue to digital converter (ADC) designed by Texas Instruments for synchronized electrocardiogram (ECG) and photoplethysmogram (PPG) signal acquisition. [35]

The AFE4900 is advertised as excellent for wearable optical bio-sensing applications due to its ability to operate with very low power requirements shown with a quoted  $30\mu A$  required for typical heart-rate monitoring scenario and  $1 < \mu A$  required when the device is powered down.

The PPG signal chain supports the use of up to four switching light-emitting diodes (LEDs) that can be switched on using the fully integrated LED driver that can even power two LEDs in parallel. It can receive, amplify and digitize the current received from up to three time multiplexed photo-diodes (PDs). These 24-Bit representations of PD current are stored in a 128 sample first in first out (FIFO) block and read out using either an  $I^2C$  or serial programming interface(SPI) interface.

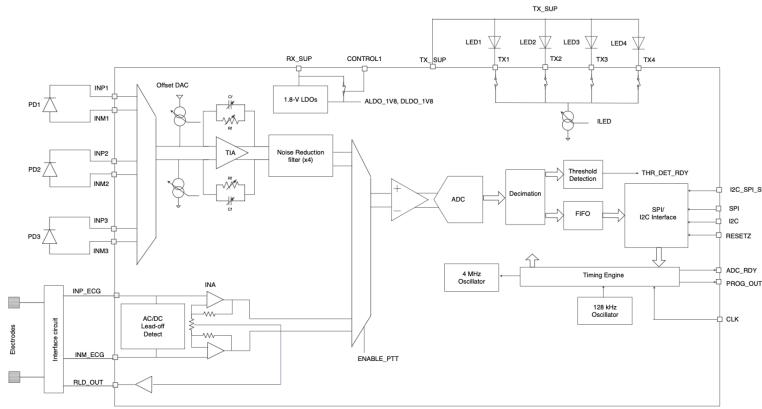


Figure 2.3: Simplified Block Diagram of AFE4900 [35]

The block diagram of the AFE is shown in Figure 2.3. For PPG signal acquisition it has the following programmable parameters.

- DC offset subtraction (Up to  $\pm 126\mu A$ ) for each LED
- Digital ambient subtraction
- Noise filtering using a filter with programmable bandwidth
- Trans-impedance gain of amplifier ( $10k\Omega$  to  $2M\Omega$ )

The AFE4900 was selected as a direct consequence of its low power compact design, its ability to power four switching LEDs and two LEDs in parallel as well as being able to receive current data from three PDs the AFE4900 was chosen. The AFE4900EVM a evaluation/development board for the AFE4900 was purchased which contained a standard sensor board to interface with the SFH7072 a multi chip PPG sensor featuring two green emitters (526nm), one red emitter (660nm) , one infrared emitter (950nm) and two detectors [36].

Stefan Karolcik, a member of the Centre for Bio-Inspired Technology then designed and produced two prototype sensor boards. The first sensor board was designed to evaluate the feasibility of the method outlined in the Samsung Patent for blood pressure estimation using PPG signals. It is simply a printed circuit board (PCB) that interfaces the SFH7050 with the AFE4900. The SFH7050 contains Green, RED, and INFRA-RED LEDs of wavelengths 525nm,660nm and 950nm respectively and one photo-diode that is sensitive to all three wavelengths [37]. By timing the switching between the LEDs and the sampling of the current from the photo-diode three PPG signals of the aforementioned wavelengths can be acquired.

The second sensor board was designed to evaluate the feasibility of hematocrit estimation using PPG signals. It comprises of two LEDs to irradiate light of wavelength 800nm and 1300nm respectively as well as two photo-diodes sensitive to regions around these two wavelengths. This initial prototype sensor board will be used to evaluate the feasibility of receiving valid PPG signals from the photo-diodes and provide a range of values for the programmable parameters of the AFE4900 that produce said PPG signals.

## 2.4 The difficulties of photoplethysmography acquisition and the need for automated quality classification

Although the advantages and utility of PPG waveforms continue to expand as research continues, the acquisition of the clear, high quality PPG signals necessary for analysis posses several difficulties. This is because PPG acquisition can be affected by numerous factors including sensing, biological and cardiovascular factors that will effect the shape of the PPG signal and limit the ability to extract the complex features required for current analysis. Table 2.1 provides a summary of said factors [38]

Factor	Examples
Sensing	Sensor geometry
	Emitted light intensity
	Sensor-skin interface
	Ambient light
	Photodiode sensitivity
Biological	Oxygen concentration
	Organ characteristics
Cardiovascular	Microcirculation volume
	Arterial blood volume
	Interstitial fluids
	Venous volume

Table 2.1: Factors altering PPG response [38]

Factors that cannot be controlled when designing a PPG acquisition device pose as the largest disadvantage to the technology. The most common of which is the voluntary or involuntary movement of the patient on which the PPG signals are being acquired. This can corrupt a PPG signal as a result of modifications to the tissue on which a PPG sensor is located due to muscle movement and the dilation of tissues. The light received by the PPG sensor will be modified due to these modifications and generate a different signal, thereby introducing an unwanted artefact [39].

Movement of a patient can also result in the displacement of the sensor relative to its original or intended location. This changes the path of light and consequently modifies the signal acquired [39], often resulting in the loss of any viable PPG signal. The initial incorrect placement of the

PPG sensor, which for some locations such as the wrist must be relatively precise [38], [40] , will also result in the inability to acquire a viable PPG signal. Added pressure applied to the device on the skin modulates the magnitude of a received signal which can result in the saturation of said signal after analog front-end processing and hence its corruption.

The anatomy of individuals, the colour of their skin as well as differences in organ sizes and the amount of fluids retained by ones tissue will all result in variation of the propagation of light through the tissue [41] introducing artefacts or even completely corrupting the ability to acquire a PPG signal for a given PPG sensor configuration.

It is possible to identify and control the artefacts introduced by these factors and hence improve signal quality by a trained individual with access to the real-time PPG signal output and open access to the configuration settings of the PPG sensor by visually classifying quality and correcting sensor placement or adjusting configuration settings accordingly, however, this is not a viable solution to the problem in the intended large scale use of the technology and efforts must be made to automate the classification of the quality of said signals such that to prompt an untrained user of their corruption which can be ignored if sporadic or corrected if continuous. The automation of such quality classification and identification of corruption will also allow for the automation of the optimisation of the acquired PPG signals by automatically adjusting sensor configurations such that to acquire optimal signals with high enough quality for the complex feature extraction required by ongoing research into the application of PPG signals.

Identifying sporadic segments of bad signal quality will not only increase the accuracy and utility of said segments in the development of any models requiring PPG signal segments as input to predict vital sign, it equally allows for a reduction in the number of false alarms that result from the segments inherent corruption. False alarms as a result of artefact corrupted signals were shown to lead to the phenomenon of "alarm fatigue" whereby nursing staff become desensitized to and ultimately ignore alerts [42], [43]. This phenomenon would prove detrimental when attempting to develop algorithms for automated alert detection or prediction diagnosis. The Joint Commission, the body that accredits U.S healthcare, highlighted the problem of alarm fatigue [42], [44] and the ECRI Institute continuously lists alarm hazards in their annual "Top 10 Health Technology Hazards" [42], [45].

## 2.5 Existing 'lightweight' approaches to PPG signal quality evaluation

As outlined in section 2.4 although the use of PPG signals has recently experienced a drastic increase in its use, the need for real-time PPG signal quality evaluation tools stems back all the way to PPG technology's inception. There have hence been a multitude of attempts to create such tools using different proposed signal quality indices (SQIs) with varying success alongside different advantages and limitations.

### 2.5.1 SQIs using electrocardiogram signals as reference

A popular and robust method for measuring PPG signal quality from the literature are studies conducted in an attempt to estimate PPG signal quality by comparing the acquired PPG signal segment to a synchronised reference electrocardiogram (ECG) signal [46]–[48]. The work conducted by Pflugradt et al in [48] presents an example of one such study, where a novel online algorithm using stochastic gradient ascent (SGA) to compare the estimated eigen-vectors of ppg signal segments was proposed to estimate the quality of PPG signals named OPRA summarised in Figure 2.4.

By using an online algorithm which self-calibrates from an initial reference period, Pflugradt et al eliminate the need for any parameters to be set specific to the devices used for acquisition and allow for the real-time classification of beats. Furthermore, by using SGA to estimate eigen vectors as appose to principle component analysis (PCA), Pflugradt et al reduce the computational complexity as well as increase the influence of individual beats in their corresponding SQI [48].

A limitation of this approach is the requirement of calibration time in which one assumes that the incoming beats can be identified and are undisturbed. Pflugradt et al claim that the convergence required for such calibration will not occur unless initially provided with beats that are undisturbed and with similar shape. The absence of convergence could possibly be used to prompt users of the likely incorrect placement or initial configuration of the PPG sensor used,

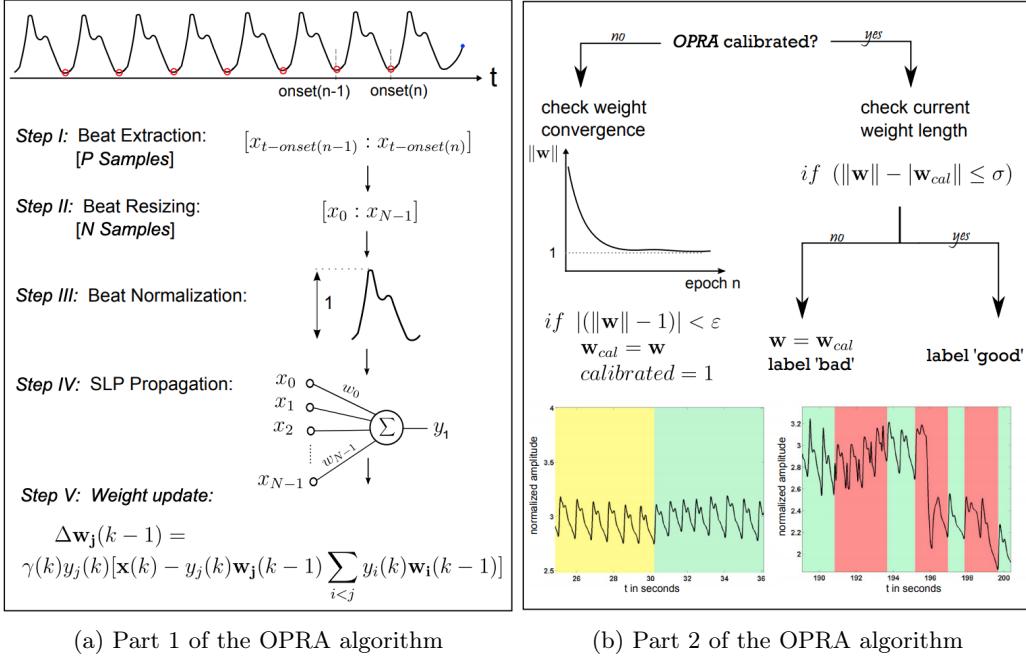


Figure 2.4: Diagram illustrating the OPRA algorithm [48]

however, no tests were conducted to support such a statement and a limit on calibration time was not provided to allow for such a utility. The main limitation of this approach is the requirement of a synchronous ECG signal to be captured such that to provide the algorithm with a ECG derived heart rate used as input to the classifier. By requiring a synchronous ECG signal one limits the potential application of the algorithm to devices in which ECG signals can be reliable captured at all times.

Although ECG technology also continues to develop towards wearable application [49], the cost of said technology remains high and the offline ability to acquire the noiseless signals required to be used as reference remains in the early stages of development

### 2.5.2 PPG independent SQIs using a dynamic morphological template as reference

As a result of the aforementioned limitations of ECG dependent PPG SQIs, several attempts have been made to develop techniques that only require PPG signals as input.

Sun et Al propose an approach that extracts four morphological and temporal features from a ppg signal segment, used as states of a Kalman filter to adaptively accept or reject signals based on their quality. The algorithm is summarised in Figure 2.5 [50]. The approach classified signals in one of four different classes with increasing signal quality as oppose to just acceptable or unfit. This extends the utility of their algorithm in comparison to others in the literature as it is capable of not only classifying signals as acceptable or unfit but also to identify signal samples which are of higher quality where specific morphological features can be identified visually. The limitation of their approach is once more the requirement of a calibration window in which it is assumed that clear PPG signals are being acquired. For said approach no claims were made as to the behaviour of the algorithm if this assumption is broken and hence the algorithm cannot be utilised when the initial placement or configuration of the sensor used to acquire the PPG signal segment is incorrect.

The requirement of a calibration window for subsequent comparison or for automatically generating thresholds is common to many other approaches for independent PPG SQI [51]–[54]. Li et Al's approach [51] to classification is performed using dynamic time warping to shape detected beats into a common template which is then compared to a reference created during calibration and Sabeti et Al's approach [52] consists of extracting a plethora of additional normalised morphological features which are fed into four different classifiers trained to predict their class. Both approaches allow for a variation of the shape and frequency of beats detected and hence increase the universality of the proposed algorithms, however, their clinical applications remain limited by

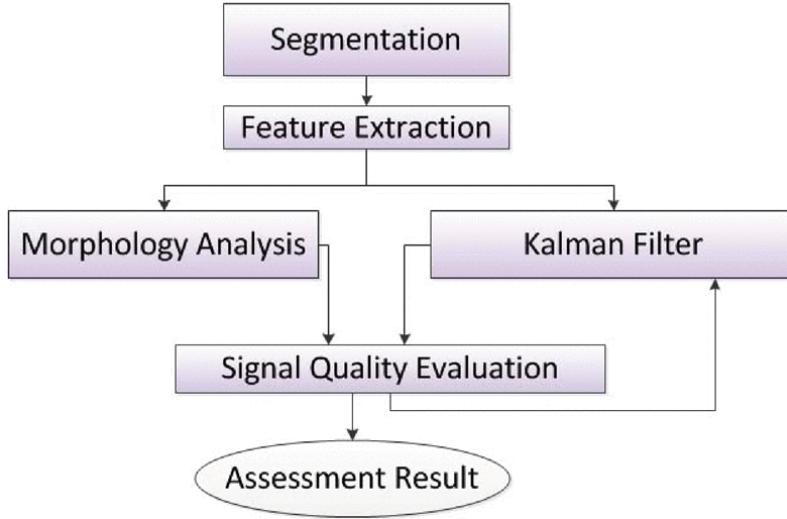


Figure 2.5: Diagram illustrating Sun et al's proposed algorithm [50]

the need of initial calibration.

### 2.5.3 SQIs for independent real-time application with no calibration

SQI indices for independent PPG classification in real-time and requiring no initial calibration have also been investigated with varying success.

Pereira et al's proposed algorithm in [55] is an example of one such approach applied in a clinical setting where the algorithm was designed to account for and hence not discriminate against signals with irregular intervals between peaks. When accurate these signals are key indications that a patient may be suffering from Atrial Fibrillation (AF) and hence should not be disregarded as this can defeat the purpose of continuous monitoring. A total of 42 different temporal-domain and spectral-domain features were extracted from each of the 30 second segments and the features extracted from a subset of all the 30 second segments were used to train three different classifiers: a support vector machine (SVM), Decision Tree classifier (DT) and a K nearest neighbour classifier (KNN). Results showed that the SVM performed best on the training data and proved that previous attempts at classifying PPG signal quality regularly considered signals with irregular intervals between peaks as unfit signal samples where as their proposed technique did not. There was no attempt to investigate the feature importance's of the 42 extracted features and hence the algorithm requires all 42 features to be extracted for each segment. Segment length was also not investigated and restricted to 30 second segments which in real-time application corresponds to a unrealistically slow refresh rate and low resolution infeasible for any real-time applications of the algorithm.

Elgendi et Al in [56] conducted a review of eight different temporal-domain and spectral-domain features with the intention of designing a 1 dimensional real-time three class classifier for PPG signal segments and identified the skewness of signal segments to be a feasible SQI for signal quality classification. The features were evaluated on their segmentation ability using a Man-Whitney U test as well as their ability to classify a dataset of PPG signal segments which they acquired and annotated themselves using an SVM classifier using a Support Vector Machine Classifier. Elgendi et Al used a three class approach to manual signal annotation allowing for the identification of signal samples which are of higher quality where specific morphological features can be identified visually. Elgendi et Al found a signal samples' skewness score to best distinguish between the aforementioned excellent, acceptable and unfit signal segments. After identifying skewness a subsequent investigation was conducted on the length of a signal sample required for accurate classifications and Elgendi et Al found that as little two seconds is sufficient and three seconds was optimal. The findings of [56] propose a three class classifier which is considerably computationally inexpensive, requiring the extraction of a single feature and a relatively fast refresh rate and fast resolution making an algorithm developed using skewness excellent for real-time applications. These findings were not validated to generalise to any PPG device however and most notably omitted the investigation of a signal sample's skewness ability to distinguish class C signals

from that of class A and B collectively. Moreover, although the use of a single SQI to classify PPG signal samples was found to be sufficient, no investigation was conducted into whether this is in fact optimal by considering a conjunction of signal SQIs and the additional computational time required.

In [57] Pradhan conducted a similar investigation, developing a classifier for the PPG signal classification of 10 second segments without calibration. 72 features were extracted for the 10 second segments and were evaluated in terms of their feature importance's when classifying the PPG signal segments, after which a list of 9 features were selected and used in a random forest classifier to classify signal segments into one of five classes corresponding to the percentage of the segment for which clear pulses with discernible peaks were identifiable. Although such an approach highlighted other features found to accurately segment samples of different classes this new classification approach ignores the ability of SQIs to identify signal segments containing beats of higher quality. Furthermore, no investigation was made on the computational intensity of the extraction of the features reviewed this omission in the investigation as well as the increased sample length limit the potential application of Pradhan's findings.

Another interesting method, proposed by Vadrevu et al in [58], determined the quality of a PPG segment using four hierarchical decisions where at each decision a feature is extracted and compared to empirically defined thresholds to finally classify the inputted six second PPG segment as either acceptable or unacceptable. Of the algorithms requiring no calibration, the algorithm proposed by Vadrevu et al was found to have the highest performance and was implemented and tested for real-time application against several datasets using different PPG signal acquisition devices. The six second input length required by the algorithm is three times as long as Elgendi et al found to be necessary for accurate classification [56] and limits the potential real time application of the algorithm. The two-class output of the proposed algorithm also limits its potential application when attempting to both identify unfit signal samples as well as higher quality signal samples. Finally the dependence on empirically defined thresholds for the four decisions required to classify signal samples and the omission of a clear annotation process used to label the signal samples and hence evaluate the performance of the algorithm further reduces the potential applicability of Vadrevu et al's algorithm for clinical application.

## 2.6 Background of proposed approach

### 2.6.1 The pre-processing of photoplethysmogram signals

The importance of pre-processing PPG signals such that to mitigate the effect that different PPG acquisition devices as well as different patients will have on any developed classification algorithm was highlighted in the literature to be of paramount importance [48], [52], [59]. A review of current approaches to the pre-processing of PPG signals was therefore conducted and the background of the selected approach is outlined as per below.

#### Re-sampling

Re-sampling of a PPG signal sample was identified in [48] to be important when attempting to classify signals acquired at different sampling frequencies and will allow for all signals acquired to be the same length, mitigating the bias that different length inputs will introduce to any potential SQIs utilised.

When re-sampling a signal the input frequency and desired output frequency will determine whether up-sampling and down-sampling are performed.

When down-sampling one is attempting to reduce the sampling frequency of the inputted signal. This will reduce the available bandwidth of the outputted signal and can hence produce aliasing effects that distort the original signal. To mitigate this effect a signal is first passed through a low pass filter before decimation can be performed to down-sample the signal. [60]

When up-sampling one is attempting to increase the sampling frequency of the inputted signal. To do so values in between the currently sampled values are interpolated often using a low pass filter to do so. [60]

In an effort to efficiently and effectively conduct such re-sampling all such operations are often conducted in the frequency domain using a Discrete Fourier Transform (DFT) when the inputted

signal can be assumed to be periodic. A discrete inputted signal to be re-sampled is first transformed into the frequency domain where upsampling, downsampling and filtration operations can be conducted before returning to the temporal domain using a Inverse Discrete Fourier Transform (IDFT).

Bluestein's algorithm [61] allows for the computationally efficient computation of the DFT of an inputted signal with a complexity of  $O(n \log_2(n))$ . The algorithm is hence an example of a method to compute the Fast Fourier Transform (FFT) of a signal and can be utilised efficiently for signals of moderate sample length.

## Filtering

As PPG signals from different datasets will have varying levels of filtration. PPG signal quality classification algorithms from the literature will often pass any data through a filter before any analysis [48], [52], [58], [59]. The importance of filtration is amplified for signals acquired and processed in real-time which will often contain high frequency artefacts present within the signal as a result of the acquisition process such as 50Hz frequency as a result of mains interference. To account for this imbalance inputted PPG signals from all datasets are often passed through a band-pass Butterworth filter before feature extraction and analysis. Figure 2.6 shows the frequency response of a nth order low-pass Butterworth filter found to have monotonic response within both the pass-band and stop-band. [62]

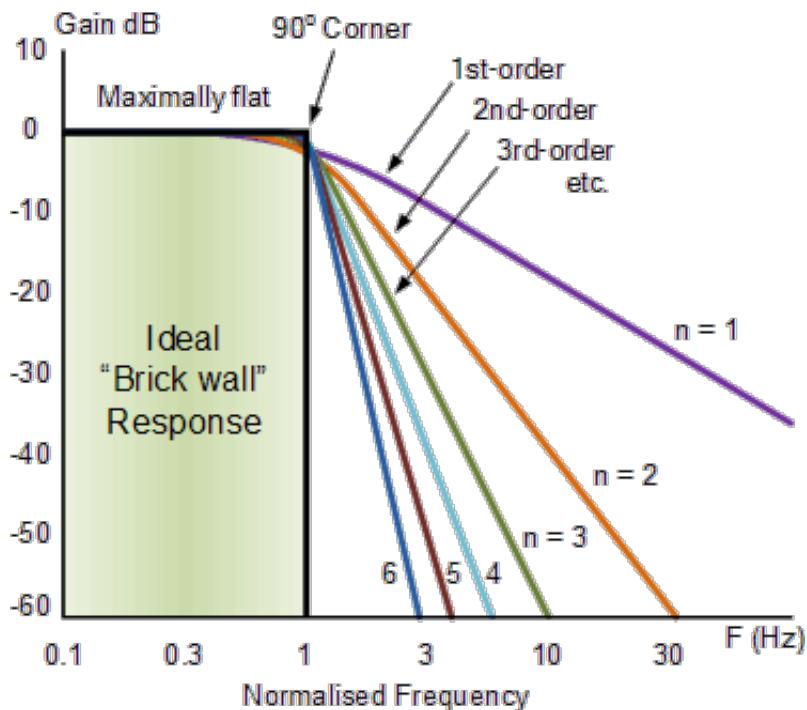


Figure 2.6: Ideal frequency response for a butterworth filter [62]

A band-pass filter is simply a low pass and high pass filter cascaded together. The order of a filter ( $n$ ) increases the analog complexity of the filter to allow for a sharper transition between the cut-off frequencies and the stop band frequencies. A digital approximation of a Butterworth filter can be implemented by designing a digital Infinite Input Response (IIR) filter with the same frequency response as that shown in Figure 2.6 for both a low-pass and high-pass filter cascaded together [63]. The simplicity of this approximation as well as the monotonic nature of the filter is why such filters are often implemented for digital PPG prepossessing. As the desired frequency components required to represent a ppg signal are at low frequencies corresponding to the low frequencies of volumetric changes in arterial blood, when filtering a ppg signal to remove high-frequency components, one does not require a small transition band between the cut off frequencies and stop band frequencies allowing for a lower order and further justifying the selection of a Butterworth filter for PPG filtration.

## Scaling and detrending

The scaling of PPG signal segments was also found in the literature to be paramount when attempting to classify PPG signals with varying amplitudes from different acquisition devices [48], [50]–[52]. Scaling a PPG signal segment can be easily implemented by finding the max and min value within a PPG signal segment. Setting said values to be equal to 1 and 0 and setting all values within the max and min values such that:

$$\text{Scaled} = \frac{\text{Raw} - \text{Min}}{\text{Max} - \text{Min}}$$

To allow for the proper implementation of some of the SQIs proposed in the literature the outputted signal is centered to have a mean centered at zero such that the equation now reads:

$$\text{Scaled} = \frac{\text{Raw} - \text{Min}}{\text{Max} - \text{Min}} - \frac{1}{2}$$

### 2.6.2 Annotation method

In [56] Elgendi et al propose a novel criteria for manual PPG signal segment annotation as per Figure 2.7.

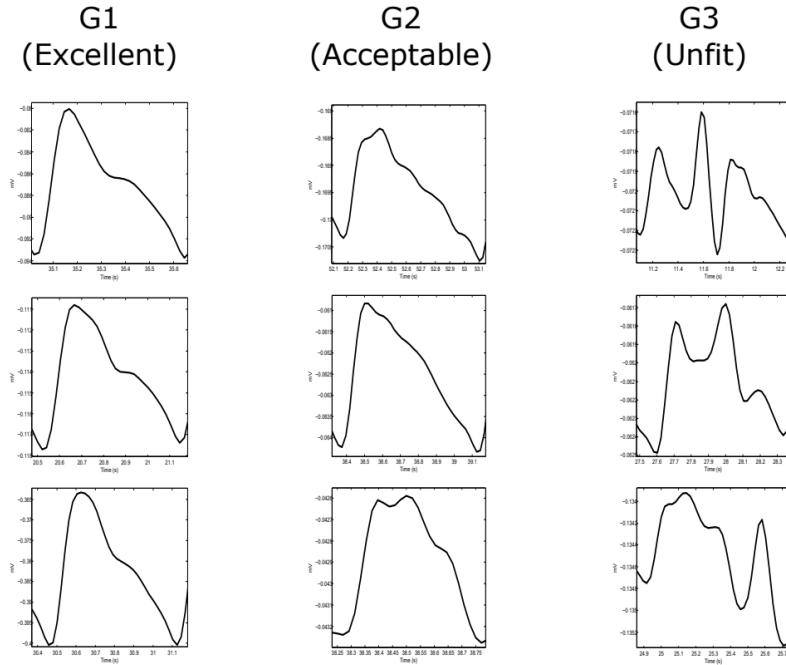


Figure 2.7: Three categories defined for PPG quality [56]

Annotators annotated PPG signal segments into one of three classes:

- Group 1 (G1) corresponds to “excellent” for diagnosis. This group only includes PPG signal segments where the systolic and diastolic waves are salient.
- Group 2 (G2) corresponds to “acceptable” for diagnosis. This group only includes PPG signal segments where the systolic and diastolic waves are not salient but where HR can be determined
- Group 3 (G3) corresponds to “unfit” for diagnosis. This group only includes noisy PPG signals where HR cannot be determined and the systolic and diastolic waves cannot be accurately distinguished.

The decision of the class of a PPG signal segment was hence decided by finding the most dominant beat morphology quality within the signal segment.

### 2.6.3 Features Extracted

#### Skewness $S_{SQI}$ as a signal quality index

Skewness is a measure of the symmetry or asymmetry of a probability distribution that was first proposed as a PPG SQI in [64] before a review by Elgendi et al found it to be an optimal SQI index for differentiating higher quality signals from lesser quality signals [56]. Skewness is defined as:

$$S_{SQI} = \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \bar{\mu}_x}{\sigma} \right]^3$$

Where  $\bar{\mu}_x$  and  $\sigma$  are the empirical estimate of the mean and standard deviation of  $x_i$  respectively and  $N$  is the number of samples in the PPG signal.

#### Kurtosis $K_{SQI}$ as a signal quality index

Kurtosis is a statistical measure used to measure the distribution of inputted data around its inherent mean and was found by Selvaraj et al to be a utile PPG SQI [65]. It allows one to understand the extent to which a distribution has a heavy tail and is sharp or a light tail and flat distribution relative to the normal distribution. Kurtosis can be calculated such that:

$$K_{SQI} = \frac{1}{N} \sum_{i=1}^N \left( x_i - \frac{\bar{\mu}_x}{\sigma} \right)^4$$

Where  $\bar{\mu}_x$  and  $\sigma$  are the empirical estimate of the mean and standard deviation of  $x_i$ , respectively and  $N$  is the number of samples in the PPG signal.

#### Shannon-Entropy $E_{SQI}$ as a signal quality index

Shannon-Entropy was also found by Selvaraj et al to be a good PPG SQI [65]. Shannon-Entropy quantifies the extent to which a probability density function (PDF) of the inputted signal differs from a uniform distribution, providing a quantitative measure of the uncertainty present in the signal [66] which is defined [67] as:

$$E_{SQI} = - \sum_{n=1}^N x[n]^2 \ln(x[n]^2)$$

Where x signal is the raw PPG signal and N is the number of data points.

#### Zero crossing rate $Z_{SQI}$ as a signal quality index

The Zero crossing rate is defined as the rate of sign-changes in an inputted signal, that is, the rate at which the signal changes from positive to negative or back and is defined as [68]:

$$Z_{SQI} = \frac{1}{N} \sum_{n=1}^N I\{x < 0\}$$

Where x is the inputted PPG signal of length N, and I, the indicator function  $I\{A\}$ , is 1 if its argument A is true, and 0 otherwise [56].

#### Signal-to-Noise Ratio $N_{SQI}$ as a signal quality index

The Signal-to-Noise Ratio (SNR) is a metric used for in a plethora of scientific and engineering fields that compares level of a desired signal to the level of background noise. There are several ways to define a SNR [69], however, Elgendi et Al propose its use as a PPG SQI when defined as the ratio of signal variance to the noise variance as follows [56]:

$$N_{SQI} = \frac{\sigma_{signal}^2}{\sigma_{noise}^2}$$

Where  $\sigma_{signal}$  is defined by Elgendi et al as the standard deviation of the absolute value of the inputted PPG signal (x) and  $\sigma_{noise}$  is the standard deviation of the x signal.

### Matching of multiple systolic wave detection algorithms $M_{SQI}$ as a signal quality index

Elgendi found that different PPG algorithms are sensitive to different types of noise in [70] and hence in [56] Elgendi et al proposed that a comparison of how accurately multiple PPG systolic wave detectors identify peaks (such as a beat or noise artifact) could provide a good PPG SQI. In [56] two well-known systolic wave detection algorithms were used. One is based on first derivative with adaptive thresholds to be set by the user [71], and the other is based on local maxima and minima [72]. These algorithms were referred as Bing's and Billauer's algorithms by Elgendi et al. Bing's and Billauer's algorithms were selected by Elgendi et al as they were both easy to implement and approached the PPG signal from different perspective [71].  $M_{SQI}$  was defined as follows:

$$M_{SQI} = \frac{S_{Bing} \cap S_{Billauer}}{S_{Bing}}$$

Where  $S_{Bing}$  represents the systolic waves detected by Bing's algorithm, and  $S_{Billauer}$  represents the systolic waves detected by Billauer's algorithm.

For the purpose of this study the Peak detection algorithm provided as part of the scipy python library [73] was implemented instead of Bing's algorithm and referred to as Scipy. When provided with a input length l as a parameter to the algorithm the Scipy algorithm finds the local maxima within consecutive lengths of l samples within the inputted signal. This remains a different approach to that of Bing's algorithm and was selected for its ease of implementation as well as its computational speed. The  $M_{SQI}$  used as part of this study was hence defined as:

$$M_{SQI} = \frac{S_{Scipy} \cap S_{Billauer}}{S_{Scipy}}$$

Where  $S_{Scipy}$  represents the peaks detected by the Scipy algorithm.

### Relative power $R_{SQI}$ as a signal quality index

Relative power was the final SQI reviewed by Elgendi et Al for its use as a PPG SQI where the frequency domain of a PPG signal was explored to asses the PPG signal quality. Elgendi et al hypothesised that as most of the energy associated with systolic and diastolic waves is concentrated within the 1–2.25 Hz frequency band [70], the ratio of the power spectral density (PSD) in this band compared to the PSD in the overall signal 0–8 Hz [70] can be used as a PPG SQI [56], defined as follows:

$$R_{SQI} = \frac{\sum_{f=1}^{2.25} PSD}{\sum_{f=0}^8 PSD}$$

The PSD was calculated using Welch's method [56], [74].

### Auto-correlation based signal quality indices

In both [57] and [75] Pradhan et al identify and validate the use of auto-correlation based PPG SQIs inspired by [76], [77]. Pradhan et al postulated that there would be a difference in measured periodicity of PPG signal samples with varying signal quality as a result of random motion artifacts present within a PPG signal sample [57], [76], [77]. By computing the autocorrelation of a inputted PPG signal and identifying the peaks of the corresponding correlogram, one can measure the length of the period of the inputted signal as well as the strength of this measured periodicity. Four SQIs were extracted using this approach corresponding to the correlation values of the first two peaks of the correlogram as well as the time lags associated with these two peaks as per Figure 2.8.

#### 2.6.4 Man-Whitney U and Wilcoxon signed-rank tests

To measure the performance of the SQIs proposed in [56], Elgendi et al use the results of a two sided Man-Whitney U test to measure the separability of G1 signal samples from samples of both G2 and G3 using each SQI score individually. A two-sided Man-Whitney U test is a non-parametric test of the null hypothesis that one randomly selected value from one population will have value less than or greater than a randomly selected value from a second population. When the two

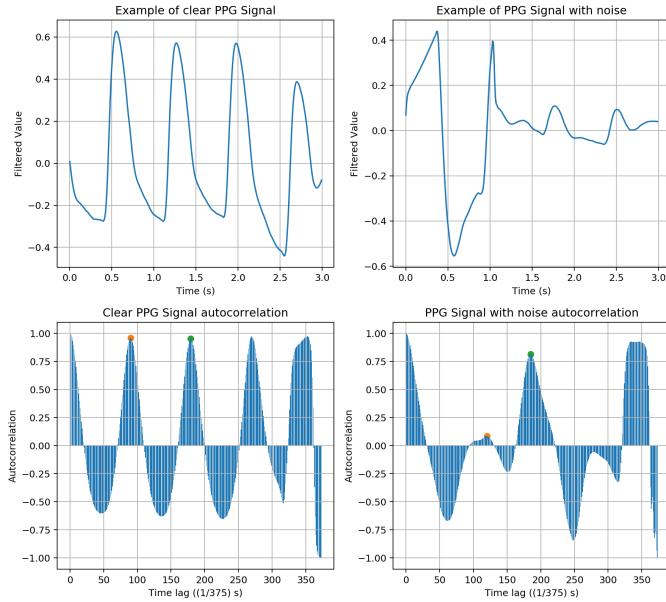


Figure 2.8: Autocorrelation SQI illustration

populations are set such that the SQI scores of certain class form one population and the SQI scores of the other two classes form the other population, a two-sided Man-Whitney U test will hence provide a indication of the variation in distribution of the different classes of PPG quality using the given SQI and hence measure the ability of said SQI to separate PPG signal samples into into their different classes. A low  $p$  value hence corresponds to a low probability of the null hypothesis and hence a large variation in the distribution of classes.

To conduct a Man-Whitney U test one must make the following assumptions [78] :

1. All the samples from both groups of classes are independent of each other.
2. The responses are ordinal.
3. Under the null hypothesis, the distributions of both populations are equal.
4. The alternative hypothesis is such that the distributions are not equal.

The most notable of said assumptions is the first one, where the all data-samples are assumed to be independent of one and other. Although one can argue that this would be the case for randomly selected PPG signal segments where the SQI score of one PPG signal segment has no relation to that of another in reality when randomly selecting PPG signal segments from a dataset of a few longer PPG recordings it is likely that PPG signals of the same quality are from the same recording and that PPG signals of differing quality may also be from the same recording at a different time. SQIs sensitive to artefacts present within a whole recording could hence be dependent on one and other. An alternative to the Man-Whitney U test is the Wilcoxon signed-rank test which is very similar to the Man-Whitney U test and tests the same null hypothesis, however, assumes that samples are in fact dependent on one and other [79], [80].

### 2.6.5 Extra random forest classifier feature importance test

Although both the Man-Whitney U and Wilcoxon signed-rank tests provide a good measure of the performance of an SQI they do not take into account the potential utility of an SQI in conjunction with another and instead only measure their independent performance. In order to measure the performance of SQIs when used in conjunction a Extra Random Forrest Tree Classifier can be fit on training examples to then provide the relative importance's of different SQIs when attempting to classify the training samples into their labeled classes [81].

A Extra Random Forrest Classifier can be created by providing randomly selected SQIs to many decision tree classifiers to form a forest structure. The individual Decision Tree Classifiers are explained in detail in section 2.6.8 and will essentially provide one with a ranked order of

optimal SQIs from the randomly selected subset that when used together and in said order best split the data into the different classes for each tree. One can then compute an estimation of the overall rank of all the SQIs reviewed to provide one with the relative importance of SQIs when used in conjunction.

### 2.6.6 The use of machine learning methods in linear classification

Once SQIs have been selected to be used for classification the optimal method of classification using said SQIs can then be investigated. The background of the classifiers reviewed in this study as well as the metrics used to evaluate their relative performance are outlined below.

### 2.6.7 Precision, Recall and the F1 Score

In order to evaluate the relative performance of a classifier one must define a set of metrics that will be used to compare the utility of different classifiers. The accuracy of a classifier is often used as said metric, however, in cases where the costs of having a miss-classified sample as either false-negative or false-positive is very high, then accuracy may not be optimal. For PPG signal quality classification if a signal is predicted to be of good quality but is in-fact of bad quality (false-positive) and hence acceptable it may lead to a false alarm which one is trying to avoid, however, if a signal is classified as bad quality but is in-fact of good quality then signals that may correspond to important and potentially dangerous changes in a patients PPG will be disregarded [82]. To measure the performance of a classifier to avoid such scenarios the following metrics can be used:

- Precision (P) is a good measure of a classifiers ability to mitigate false-positives and is defined such that:

$$P = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

- Recall (R) is a good measure of a classifiers ability to mitigate false-negatives and is defined such that:

$$R = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- F1 score provides a metric that is a balance between the Precision and Recall scores defined such that:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

The difference between a F1 Score and the accuracy of a classifier is that when class distribution is uneven the accuracy of a classifier can often ignore the classification accuracy of the minority class distribution. When the minority class is inaccurately classified by a classifier this can result in an inaccurate representation of the performance of a classifier. The F1 score avoids such inaccuracies and provides a metric of the balanced performance of a classifier per class [83]. The macro-average F1 score measures the average of the F1 scores for each of the classes available equally and was hence used to evaluate the relative performances of the different classifiers reviewed.

### 2.6.8 K-fold Cross-validation

K-fold Cross-validation is a popular method to measure the performance and generalisation of a classifier when a validation set cannot be provided [84]. A training set is randomly divided into K subsets such that in one of K iterations, K-1 subsets are used for training and the left over subset is used for validation where in each iteration a different subset is left out. Figure 2.9 illustrates this procedure when K=10.

#### Soft Margin Support Vector Machines (SVMs)

Support Vector machines (SVMs) were the most popular classification tool from the literature and were found by Pereira et al in [55] to provide the highest accuracy of the classifiers reviewed within that study.

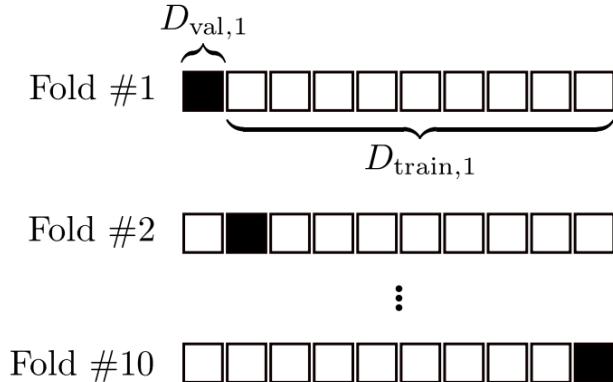


Figure 2.9: 10-fold cross-validation illustration [84]

The objective of a SVM is to find a hyper-plane in an N-dimensional space that distinctly classifies the data points where N is the number of features available to the classifier [85].

When separating two classes of data points, there are many possible hyper-planes that can be selected. The objective of a SVM however is to find a plane that maximises the distance between data points of both classes (Maximum Margin) in an attempt to predict the classes of future data points with greater confidence.

## Support vectors

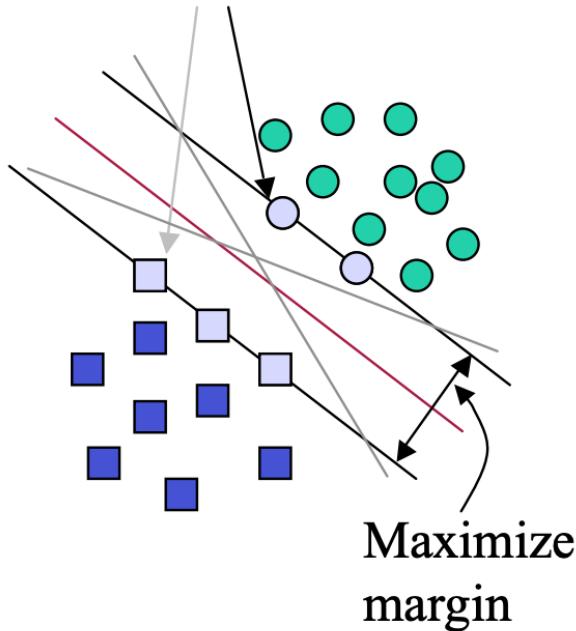


Figure 2.10: Diagram illustrating support vector machine hyper-plane formulation [85]

Support vectors are data points that are closest to the hyper-plane and hence will influence the position and orientation of the hyper-plane. A soft margin SVM is one where the total separation of classes using a hyper-plane is not required when maximising margin and the loss function that helps to maximize this margin is hinge loss and is defined as follows:

$$\min_{\omega} \lambda \|\omega\|^2 + \sum_{i=1}^N (1 - y_i \langle x_i, \omega \rangle)$$

Where  $\omega$  is a vector of coefficients for the feature space  $X$  that together form the hyper plane

used for classification such that if the scalar product of  $\omega$  and a specific sample  $x$  is  $> 0$  then said sample is classified as class 1 and class 0 otherwise.

The first term in the loss function in the above equation corresponds to the introduction of regularization parameter  $\lambda$  that balances margin maximization and loss thereby allowing for a soft margin to be found. This loss is subsequently minimised with respect to  $\omega$  to find the optimal  $\omega$  which corresponds to a hyper-plane that separates the two classes with largest possible margin for a given  $\lambda$ .

Multi-Class SVMs are often implemented using a one versus all approach where a two-class SVM is created for each class against all other classes after-which the class of a specific training sample is selected by finding the class that corresponded to the greatest margin for said training sample [86]. Kernels can be applied to allow for the non linear separation of data with minimal computation requirement [87]. Examples of two of such kernels are:

- **Polynomial kernel:**  $k(x_i \cdot x_j + 1)^d$
- **Gaussian Radial Basis Function:**  $\exp(-\gamma \|x_i - x_j\|^2)$

Non-Linear kernels however are often susceptible to overfitting, a phenomenon which occurs when a classifier achieves a high accuracy on the training data used to fit said classifier however achieves a lower accuracy on data unknown to the model as the classifier has been tailored specifically to the distribution of classes within the training data and not the general distribution of classes. A regularisation hyper-parameter ( $k$  and  $\gamma$  in the example kernels) is hence used to regularise the complexity of said kernels [88].

### Linear Classifiers using other loss functions

As outlined in the previous section a SVM is simply a linear classifier trained to minimise hinge loss with respect to the coefficients of the hyper-plane used to classify a sample. There are several other losses that can be minimised which may instead be optimal when classifying PPG signal segments using a given set of SQIs, the losses reviewed in this report are outlined below:

- **Ridge Regression** uses a loss function with  $l_2$  penalty such that [89]:

$$\lambda \|\omega\|_2^2 + \sum_{i=1}^N (1 - y_i \langle x_i, \omega_i \rangle)^2$$

- **Elastic Net Loss:** This loss function combines both  $l_1$  and  $l_2$  penalties by minimising [89]:

$$\min_w \frac{1}{2n_{\text{samples}}} \|X\omega - y\|_2^2 + \alpha\rho\|\omega\|_1 + \frac{\alpha(1-\rho)}{2}\|\omega\|_2^2$$

It is useful when features are correlated with one another where  $\alpha$  and  $\rho$  are hyper-parameters.

- **Modified Huber:** Huber loss was introduced to provide a loss function robust to outliers and a modification of this loss function is often used in classification which is simply the quadratically smoothed hinge loss such that [90]:

$$L(y, \omega, x) = \begin{cases} \text{Hinge Loss}^2, & \text{for } y(\omega \cdot x) \leq -1 \\ -4y(\omega \cdot x), & \text{otherwise} \end{cases}$$

The regularisation term is inherited from the Hinge loss classifier discussed in the previous section.

- **Logistic Regression** is a loss function that aims to maximise entropy with a loss function decided such that [89]:

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^n \log(\exp(-y_i(x_i^T \omega + c)) + 1)$$

Where  $C$  is the regularisation parameter and  $c$  is the intercept of the hyper-plane.

- **Squared Hinge Loss** this loss function simply returns the Hinge Loss squared. It is often utilised such that to increase the penalty on large errors and slightly reduce the penalty of small errors [89]. The regularisation term is also inherited from the hinge loss classifier discussed in the previous section
- **Perception Classifier** A perception classifier is simply a hinge loss classifier with no regularisation parameter where stochastic gradient descent is instead used to find an optimal  $w$  that provides an estimated minimum loss [89].

Multi-Class Linear Classifiers can be implemented using the same approach outlined for multi-class SVMs in section 2.6.8.

### Shallow Decision Tree Classifiers

Contrary to the findings in [55] Pradhan et al found that decision tree classifiers were optimal for PPG signal quality classification. A decision tree classifier is a visual and simplified algorithm for making decisions. It consists of a set of decision nodes and leafs: decision nodes specify a test that a single attribute value is subject to with one branch and sub-tree for each possible outcome of the test. Leafs are at the end of the tree indicating a class that a case will be classified as if said leaf is reached [91]. A decision tree is used to classify a case by starting at the root of the tree and then passing through decision nodes shifting attention to the sub-tree that belongs to the case's outcome for a given decision node until a leaf is reached [91].

A popular algorithm for constructing a decision tree is the C4.5 algorithm introduced in [91] by Ross Quinlan. For each decision node starting at the root of a tree the C4.5 algorithm chooses the attribute of the data ( $a_{best}$ ) that most effectively splits its set of samples into subsets resulting in the highest information gain and produces a decision node at this point that tests for  $a_{best}$ . Information gain (IG) is defined as the change in information entropy after an attribute of data has been selected to distinguish between the dataset and can be calculated as follows:

$$IG(T|a) = H(T) - H(T|a)$$

Where  $H(T)$  is the Shannon entropy of the training set [92] and  $H(T|a)$  is defined as:

$$H(T|a) = \sum_{v \in vals(a)} \frac{|S_a(v)|}{|T|} H(S_a(v))$$

Where:  $S_a(v) = \{x \in T | x_a = v\}$  : The set of training inputs T for which attribute a=T

The algorithm also contains three base cases where:

- If all the samples in the list belong to the same class, create a leaf node for the decision tree saying to choose said class.
- If none of the features provide any information gain, create a decision node higher up the tree using the expected value of the class.
- If a instance of previously-unseen class is encountered, Create a decision node higher up the tree using the expected value.

By limiting the depth of a decision tree to create a shallow decision tree one is essentially provided with a series of thresholds which when extracted may provide viable thresholds for linear classification.

# Chapter 3

## Requirements Capture

### 3.1 Signal Quality Algorithm:

After reviewing previous attempts at developing algorithms for PPG signal classification in section 2.5 the following criteria were set to serve as a guide for SQI review, classifier evaluation and algorithm development.

#### 3.1.1 Independence

As identified in section 2.5.1 the requirement of inputs not extracted from the tested ppg signal segment, such as ECG derived features, limits the potential application of any algorithm or framework that requires said inputs to applications in which a ECG signal is also attained. As wearable, reliable and continuous ECG technology is still in its infancy [49] and to allow for the wide scale application of the findings of this study, the features reviewed as part of this study, that would form the basis of any potential algorithm or framework for PPG signal classification, were limited to those that can be extracted from a PPG signal segment.

#### 3.1.2 Universality

Another limitation of several proposed algorithms for PPG signal classification as identified in section 2.5.3 was that the datasets used to review both SQIs and classifiers were often limited to PPG signals acquired from just one device. As a result it is difficult to argue that such SQIs' and classifiers' utility will extend to other devices with different ranges, amplitudes and sampling rates. It was hence required that the dataset used in this study when reviewing the utility of SQIs and classifiers for PPG signal classification is acquired from a range of different devices at different sampling rates.

#### 3.1.3 Real-time application

A range of different algorithms for PPG signal classification were reviewed in section 2.5, some of which were developed for real-time application and others assume that the entirety of the signal is available for analysis. This study focuses on reviewing SQIs and classifiers and proposing a algorithm/framework to classify PPG signal segments in real-time. The signal sample length provided as input when extracting the SQIs reviewed was therefore limited to three seconds as a result of the findings in [56] where Elgendi et al show that a three second input length was sufficient for accurate classification.

#### 3.1.4 Utility

The SQIs and classifiers reviewed in this study must be evaluated on their ability to classify signals into one of three classes : excellent, acceptable or unfit for analysis. This is such that to allow for a wide range of potential applications of any algorithm/framework developed as outlined in section 2.4 an example of which is the possible integration of one such classification algorithm into an automated calibration algorithm for the setting of initial configurations of a PPG acquisition device, to aid in the initial automated acquisition of a viable PPG signal as well the optimisation

of these configuration settings to improve the quality of any measured PPG signals and allow for the extraction of more difficult to measure features from a PPG signal such as the locations of the discrotic notch and diastolic peak within beats.

To extend the application of any developed algorithm to identify initial incorrect PPG sensor placement, SQIs and algorithms which require calibration a period in which it is assumed that the incoming PPG signal is clear to identify a template to compare incoming signal segments to, should be omitted from review.

### **3.1.5 Computational requirement**

As identified in section 2.5.3 a limitation of several proposed SQIs and algorithms for PPG signal classification is that no review of the computational requirement and intensity of said tools was conducted. This can limit the potential embedded application of such algorithms if they are of high computational complexity and hence require a substantial amount of processor time for computation. Any SQIs proposed must hence be reviewed for their computational intensity. As computation time is relative to the processor on which computation is performed no distinct thresholds were set for the limit of the computation however an emphasis must be made on limiting the computational complexity of any proposed algorithm/framework.

## **3.2 Graphic User Interface:**

To allow for the simple application of any algorithms proposed by this study or created using the framework outlined from this study a Graphic User Interface (GUI) should be developed that allows for the simple real-time application of these algorithms as well allowing for the classification of pre-recorded signals. The criteria for the development of this GUI are outlined below:

### **3.2.1 Compatibility**

To allow for the simple integration of the GUI into both existing and future studies requiring PPG signal acquisition and analysis any GUI developed must have full cross-platform capability.

### **3.2.2 Ease of Use**

The GUI developed should be relatively simple to use. Although it is intended to be used by those familiar with PPG signal acquisition its main purpose is to showcase the potential of any developed algorithm to classify PPG signal segments in real-time and hence must include a real-time visualisation of the current PPG signal segment quality as well a method to visually confirm the accuracy of an algorithms' classifications.

### **3.2.3 Utility**

To increase the utility of the GUI one must be able to both acquire and classify PPG signals in real-time as well as import and classify finite PPG signals that have been pre-recorded. It must also be capable of importing configurations for algorithms developed by others using the framework outlined in this study such that to be able to test the performance of thresholds selected by the user to classify PPG signals

### **3.2.4 Export Capabilities**

Raw and Filtered Signals attained using the GUI in as well as the annotations of the quality of PPG signal segments within the recorded signal must be exportable to allow for simple extraction and use in the further analysis of PPG signals as per the study for which they were recorded.

# Chapter 4

## Analysis and Design

### 4.1 Data Acquisition:

As mentioned in the previous chapter a key requirement in the design of the algorithm is its intended universality. This is because PPG signals are currently acquired at a range of locations, wavelengths and sample rates with varying levels of inherent filtration and scaling. These variables have an effect on the distribution of the scores reviewed as part of this project and thresholded by the algorithm and hence to claim universality the algorithm would have to be designed by considering data from a range of sources, sensors and probes to form one dataset used to develop the algorithm for signal quality classification. This section describes the data acquired for this purpose.

#### 4.1.1 The BIDMC Dataset

The BIDMC dataset [93] was acquired as a subset of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) dataset [94], a public-access intensive care unit database hosted on PhysioNet [95]. The patients in the dataset were selected from a larger cohort of patients who were admitted to medical and surgical intensive care units and had their PPG readings taken at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, USA. The dataset consists of 53 PPG recordings of 8-minute duration sampled at a sampling rate( $f_s = 125$  Hz), from 53 adult patients with median age: 64.81, age-range: 19-90+, of which 32 were female. From these 53 PPG recordings 50, 3 second samples were sampled at random without replacement from each recording providing 2650 independent samples.

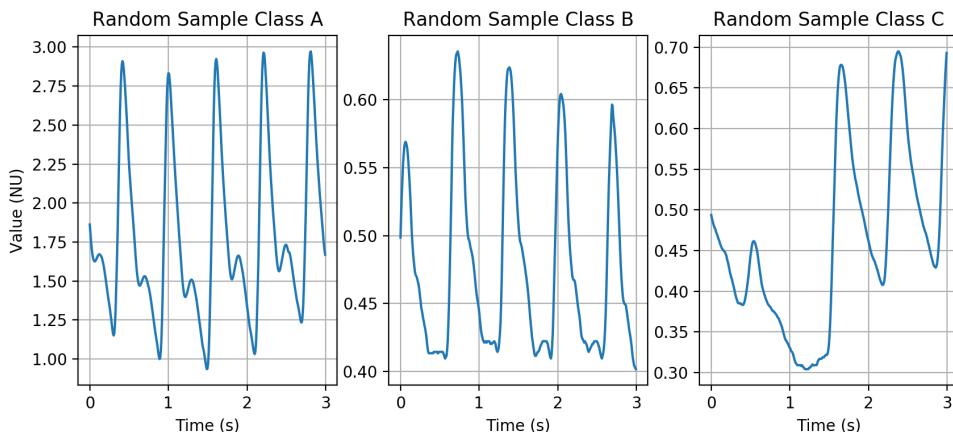


Figure 4.1: Examples of class A, B and C samples from the BIDMC dataset

Figure 4.1 shows an example of the three raw samples from the dataset of three annotated classes. The dataset provides the signal with normalised units (NU) and appears to have been filtered however the paper describing the dataset [94] provides no indication of the normalisation or filtration process. It must also be noted that the amplitude, shape and mean of these samples

vary, this suggests the use of several wavelengths, probes and devices to measure the PPG signals although this is again not explicitly mentioned. The variance in amplitude, shape and mean of these samples will nevertheless provide a general dataset of PPG samples.

#### 4.1.2 Data collected at the HTD, Vietnam

The next subset of data acquired for subsequent analysis is taken directly from the Hospital for Tropical Diseases (HTD) in Ho Chi Minh City, Vietnam where the algorithm developed as part of this project is first intended to be used. The dataset was collected using an open source pulse oximeter developed by SmartCare devices [96]. The dataset consists of 3 PPG recordings with different duration's sampled at a sampling rate( $f_s$ ) = 100 Hz, from one adult patient admitted to HTD over the course of two days. 50 samples 3 second samples were taken from each of the three recordings to provide a subset of 150 samples.

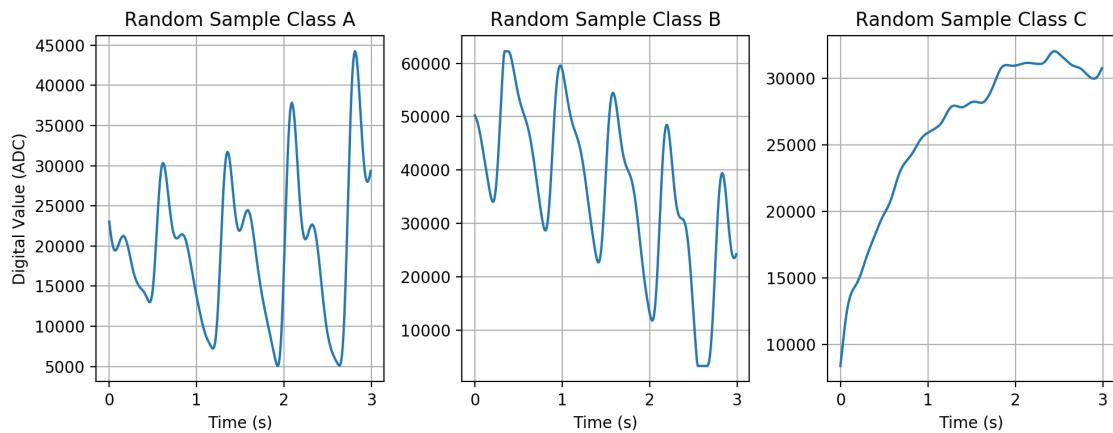


Figure 4.2: Examples of class A, B and C samples from the Vietnam HTD dataset

Figure 4.2 shows an example of the three raw samples from the dataset of three annotated classes. The dataset appears to provide the signal in its raw ADC value and appears to have been inherently filtered by the device however no documentation for the acquisition device was found hence the resolution and filtration process was unknown.

#### 4.1.3 Data collected using the AFE4900EVM module

The AFE900EVM module previously described in section 2.3 is the intended acquisition device for a future study at the HTD in Vietnam by the Centre For Bio-Inspired Technology at Imperial College. As outlined in section 2.3, PPG signals of different wavelengths will be acquired using various probes and the AFE4900EVM module. These probes are currently in development however preliminary data was collected from three different sensor configurations and placements. 331, three second samples were collected from the fingers of three different people using the SFH7072 reflective pulse oximeter explained in section 2.3 at a infrared wavelength of 950nm and a sampling rate  $f_s = 250$  Hz. A further 30, three second samples were collected from the finger of one person also using a SFH7072 reflective pulse oximeter but at two wavelengths of 950nm and 660nm with 15 samples for each at a sampling rate  $f_s = 500$  Hz. Finally another 45, three second samples were collected from the wrist of one person using the SFH7050 reflective pulse oximeter at three wavelengths of 950nm, 660nm and 525nm with 15 samples for each at a sampling rate  $f_s = 500$  Hz.

Figure 4.3 an example of the three raw samples from the AFE dataset for the three different classes. These raw samples can clearly be seen to be unfiltered and have varying amplitudes and means. In order to get a clearer picture of their varying quality a plot of the filtered, down sampled and scaled signals can be found in the same figure.

## 4.2 Data Processing:

It is clear from Figures 4.1, 4.2 and 4.3 that the raw data collected from all three databases vary immensely with different sampling rates, amplitudes and mean values as well as different artefacts

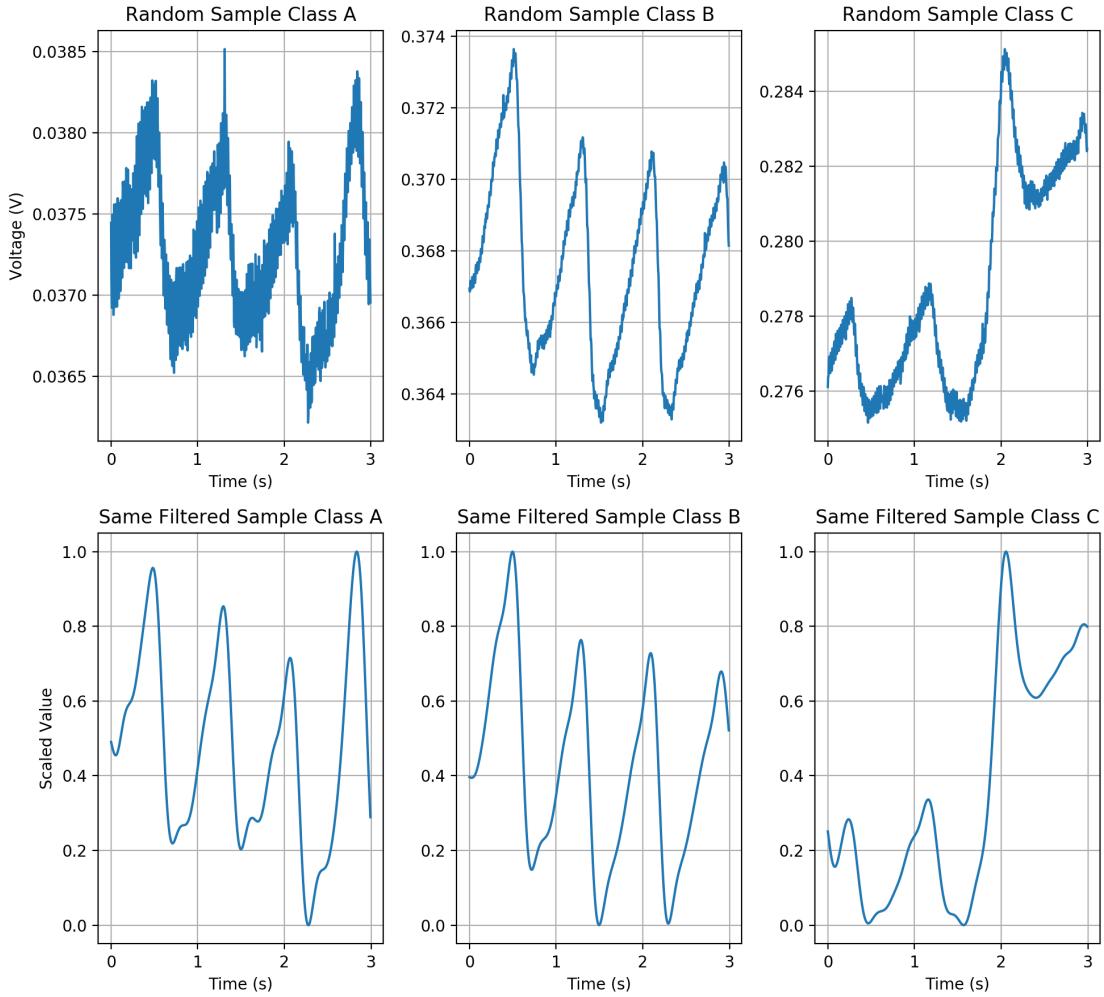


Figure 4.3: Examples of class A, B and C samples from the AFE dataset

present in the C class labeled data samples. The variance of all such characteristics of the signals is desirable when claiming universality of any algorithm developed and tested using this data, however, to be able to conduct any meaningful analysis of this data, one must develop a common pipeline for the pre-processing of said signals before the various SQI's of each signal are calculated.

This pipeline of pre-processing must be developed with the intention of real-time application as per section 3.1.3. This is the reason for why the raw signals captured as detailed in section 4.1 have been split into 3 second samples which can then be individually processed in real-time.

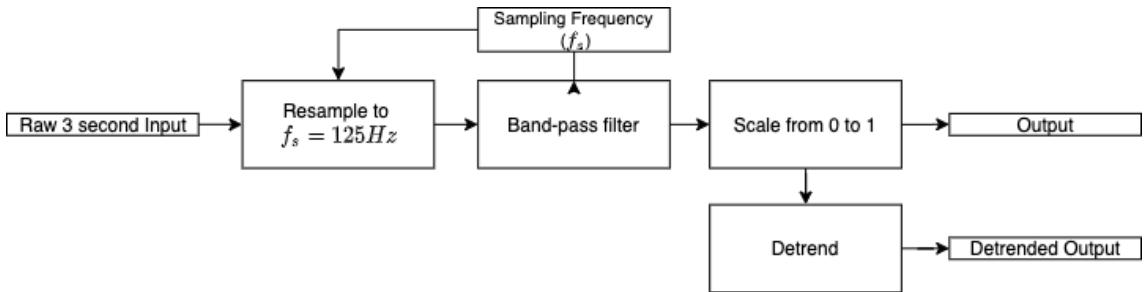


Figure 4.4: Diagram of pre-processing pipeline

The diagram illustrated in Figure 4.4 shows the pipeline that a raw 3 second data sample is processed through before the various SQI scores were measured on the output. This pipeline requires for the sampling rate of the inputted signal to be known such that the first step in processing is to re-sample the signal input to a sampling rate ( $f_s$ ) = 125 Hz according to the re-sampling method explained in 2.6.1. If the sampling rate of the signal sample is already 125 Hz

it remains unchanged and passes through to the next step. This base sampling rate was chosen as not only is it the standard sampling rate for any continuous physiological signal in any of three large MIMIC databases, including the subset utilised for the development of the algorithm, it is also the default sampling rate for the AFE4900EVM and all other sampling rates available on the device are a multiple of 125.

The next step in the pre-processing pipeline is once more dependent on the knowledge of the sampling rate to pre-initialise a band pass digital butter-worth filter with a high cut off frequency ( $f_{highc}$ ) = 30 Hz and a low cut off frequency ( $f_{lowc}$ ) = 0.5 Hz an order = 6 which the output of the re-sampling stage is subsequently passed through. This method of filtration of PPG signals is explained in section 2.6.1 and was chosen as a simple filter to pass any data sample through regardless of if or how it has previously been filtered before passed to the algorithm. Admittedly this filter was decided upon for its performance in handling the raw data acquired and filtered by the AFE4900EVM as it is this device that is intended to be used in future studies however as the filter is part of the algorithm intended to be universal for any PPG signal such a simple filter also allows for the algorithm to be sensitive to any prior filtration or absence in filtration of an inputted signal.

Finally the sample is then scaled to zero and the maximum value of the sample is scaled to 1. As clear in Figure 4.4 a scaled and detrended version of the signal is also outputted where the detrending of the signal is outlined in section 2.6.1 performed for SQI scores that were found to have been corrupted by variations in the DC value of the signal. Examples of the effects of this pre-processing on the raw data can be found in Figure A.1 in the Appendix.

### 4.3 Data annotation:

All 3 second samples collected were then pre-processed individually as per the pipeline outlined in the previous section after which all data was annotated following the procedure proposed by Elgendi et Al [56] and described in 2.6.2. Excellent signal segments named group one in Figure 2.7 will hereon be referred to as class A signals, acceptable signals named group two in Figure 2.7 will be referred to as class B signals and unfit signals named group three in Figure 2.7 will be referred to as class C signals. If any artefact was present within the signal sample which would visibly corrupt any diagnosis using the signal sample the aforementioned sample was labeled C regardless of the rest of the majority sample quality. This is as a result of the short sample length and the desire to pick up on any artefact within the signal that would corrupt diagnosis as outlined in section 3.1.4. No penalty was awarded to signal samples where the distance between peaks was uneven such that to avoid classifying signals detecting Atrial Fibrillation as unfit as per [55]. Examples of such annotations can be found in Figures 4.1, 4.2 ,4.3.

Dataset	# Class A Samples	# Class B Samples	# Class C Samples
<b>MIMIC Subset</b>	462	1824	364
<b>Vietnam Subset</b>	26	109	15
<b>AFE Subset</b>	3	206	257
<b>Total</b>	<b>491</b>	<b>2139</b>	<b>637</b>

Table 4.1: Class distribution of acquired data

It is clear from Table 4.1 that there is an uneven representation of the three classes within our total dataset which contains significantly more class B signals than that of class A or C. The explanation for this uneven distribution comes from the knowledge of the real global distribution of such signals when physically acquired. That is to say that in practical signal acquisition one is much more likely to acquire signals of quality B than that of quality A and datasets such as the MIMIC dataset often omit signals from their dataset if they are predominantly of class C. The distribution of the classes within the dataset can hence be assumed to be the same as the global distribution of PPG signal segments.

## 4.4 SQI Analysis:

### 4.4.1 Train, Validation and Test Split

Before completing any analysis on the ability of a given SQI index in segmenting the data it must be noted that all such tests were performed on a training set of the entire acquired dataset. This training subset was formed by shuffling each acquired subset randomly before selecting 60% of the subset for training and 40% for testing. A comparatively large testing set was utilised as a result of the poor class distribution of the acquired dataset in an effort to ensure that there were enough classes within the subset to ensure the classes feature distribution within the subset was reflective of the full dataset. All three subsets outlined in section 4.1 were incorporated equally into the training and testing sets such that to preserve the universality of the training data.

### 4.4.2 SQIs selected

After completing both the pre-processing and the annotation of the data acquired an evaluation of the signal quality indices proposed in the literature to classify PPG signals can now be completed.

As outlined in section 3.1.3 the algorithm is required to measure signal quality indices that can be measured using only a 3 second sample of a PPG signal as input with no prior information on the sample received as input and no reference signal. This decision has been made to ensure the versatility of the algorithm developed and ensure that readings from sensors that have been placed incorrectly from the start of any measurement can be correctly classified as noisy signals.

Given these strict requirements 26 SQIs proposed in the literature and explained in 2.5 were analysed: skewness, kurtosis, entropy, zeros, MSQ, SNR and relative power were all computed on a full data sample as per [56]. The pulse wise skewness, kurtosis, entropy, SNR, and relative power were also computed by identifying pulse segments using the Billauer peak detector explained in 2.6.3 before passing each segment to be measured for the aforementioned SQIs. As per [57] The mean, median and standard deviation of these pulse wise SQIs were calculated per sample and were subsequently used as SQIs of the sample. The lag location of the first and second peak of a samples autocorrelation were also found and used as two separate SQIs as well as the value of the peak at said locations.

These SQIs were selected as a result of their reported success in real time classification by Elgendi et Al in [56] as well as by Pradhan et Al in [57]. Although both studies conclude on a selection of features that were found to be optimal for their data selection and specific method of annotation and data length, all SQIs that were measured using only the PPG signal were re-evaluated as part of this project. This is because both papers draw their conclusions from locally acquired datasets and both their class annotation techniques and data sample length were different to those proposed in this study.

SQI	SQI #	SQI	SQI #	SQI	SQI #
Skewness	0	mean_Entropy	9	std_Kurtosis	18
Kurtosis	1	mean_SNR	10	std_Entropy	19
Entropy	2	mean_Relative_Power	11	std_SNR	20
Zeros	3	median_Skewness	12	std_Relative_Power	21
MSQ	4	median_Kurtosis	13	AC_loc1	22
SNR	5	median_Entropy	14	AC_peak1	23
Relative_Power	6	median_SNR	15	AC_loc2	24
mean_Skewness	7	median_Relative_Power	16	AC_peak2	25
mean_Kurtosis	8	std_Skewness	17		

Table 4.2: SQI number table

A notable SQI omitted from review is the perfusion score of a signal sample. It is considered the gold standard of PPG signal quality assessment by several pulse oximetry manufacturers, such as Philips, Nellcor-Medtronic, and Masimo [56], [97]–[100] and is defined as:

$$P_{SQI} = \frac{y_{max} - y_{min}}{\bar{x}} \times 100$$

Where  $P_{SQI}$  is the perfusion index,  $\bar{x}$  is the statistical mean of the raw PPG signal and  $y$  is the filtered PPG signal [56]. Perfusion scores were omitted from analysis as the raw signal segments required for the calculation of perfusion score were not available from the MIMIC and Vietnam data subsets as outlined in section 4.1.

The range of perfusion scores of the other samples within the dataset found to differ immensely classification algorithms. This variation in the range of perfusion scores is as a result of the analog front end signal processing conducted on the PPG acquisition device used to record the samples, as outlined in section 2.3.3 the AFE4900 has the ability to introduce a current offset to the output of a photo-detector, this ability is common to most PPG acquisition devices and changes the mean of the raw outputted signal as desired by the designer of the acquisition device without necessarily changing the signal quality. Without the knowledge of the magnitude of this change which will differ from device to device depending upon the intended use of the device perfusion scores are hence unfit as a SQI for universal threshold classification.

#### 4.4.3 MSQ score implementation

As explained in section 2.6.3 parameters are required to be selected by both Billauer's peak finding algorithm and Scipy's peak finding algorithm such that to compute a MSQ score. The two parameters; delta for Billauer's and length (distance) for Scipy's were selected by conducting a two dimensional grid search for the optimal parameters using the training data.

The grid search was used to develop the MSQ score function independent of the other scores implemented and potentially used by the algorithm. This selection can be conducted independently as parameters that are selected to maximise data segmentation in one dimension , will correspond to a maximum segmentation on that axis of further dimensions. This is because the computation of the scores used are independent of one and hence the distribution of the PPG samples on the MSQ dimension will have no effect on the distribution on other dimensions.

This grid search used the p value of the two-sided Man Whitney U test explained in section 2.6.4 on the segmentation of classes AB vs C and A vs BC as the score aimed to be minimised by parameter selection.

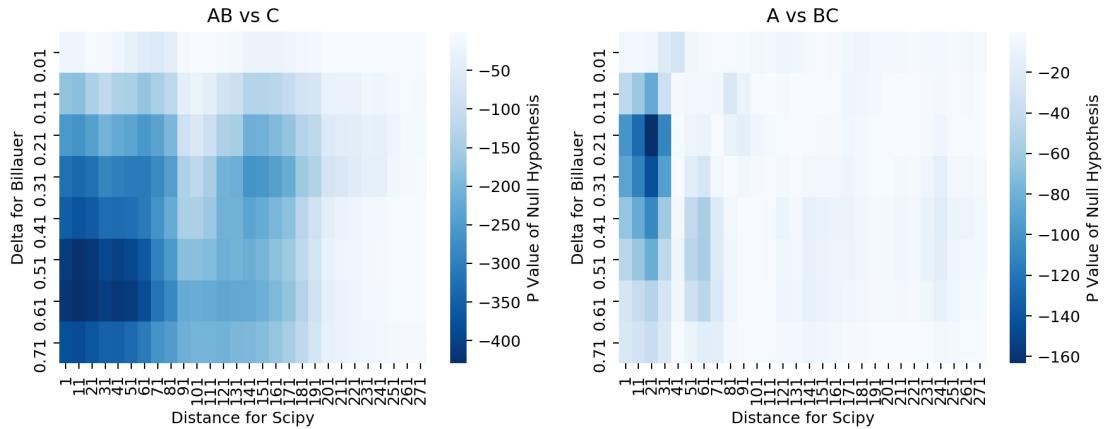


Figure 4.5: Two dimensional Grid Search for Optimal MSQ parameter selection

Figure 4.5 shows the results of this grid search where it must be noted that parameters that corresponded to optimal AB vs C segmentation did not coincide with those that corresponded to optimal A vs BC segmentation. As a result the delta value ( $\delta$ ) was selected to be such that  $\delta = 0.4$  and the optimal distance between peaks required by the Scipy peak finder ( $d$ ) was selected to be such that  $d = 21$ . These values were selected empirically and deemed to be an optimal trade-off in segmentation ability for the two classification problems.

#### 4.4.4 Computation time

Before performing any analysis on the effectiveness of the selected SQIs in segmenting the different classes of the acquired data one must first perform an analysis of the relative computational time required to compute said scores. This analysis was proven necessary such that to meet the

requirements outlined in section 3.1.5 where computation time and intensity were identified as key regularisation factors in the development of the algorithm.

To calculate the computation time required to measure a given SQI the standard python package "timeit" [101] was used as it avoids a number of common traps for measuring execution times [102]. It must once again be noted that these execution times are relative to the processor on which they were computed and will vary, however, an analysis of the relative computation time of SQIs can still be conducted. The computation time for a given SQI was measured as the time it required for both pre-processing and computation as some SQIs required the signal to be additionally detrended before computation and this extra time required was desired to reflect upon its computation time.

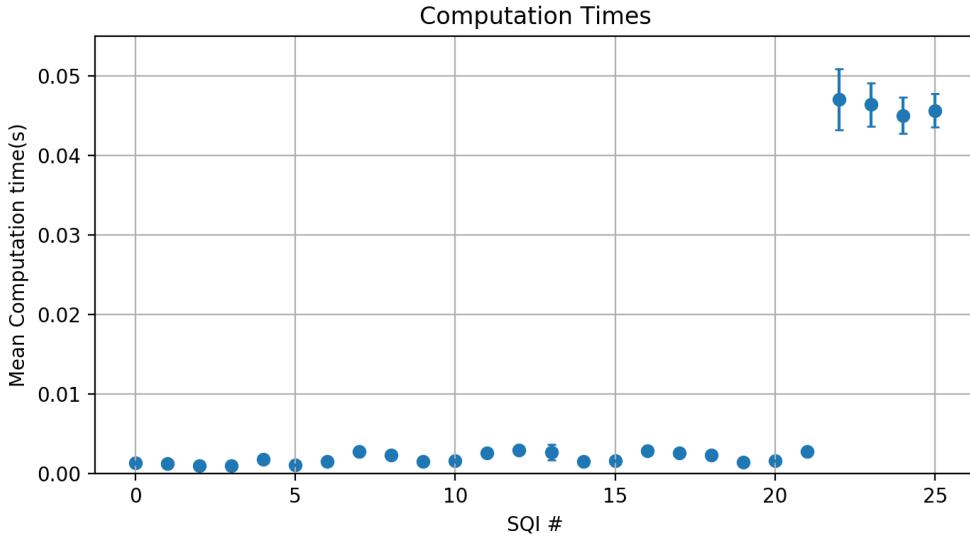


Figure 4.6: Computation time of all computed scores

It is clear from Figure 4.6 that the relative computation time between all non autocorrelation based SQIs were similar indicating that the bulk of computation time was spent pre-processing the signal. It must be noted that the computation time required to measure the per pulse SQIs will increase linearly with the increase in pulse of the patient on which the data sample is taken from as a higher pulse corresponds to more segments required to measure the SQI per pulse in a fixed data sample.

A comparison of the time required to compute all non autocorrelation based scores with the computation times of the autocorrelation based scores in Figure 4.6 shows a sharp increase in the computation time required to measure all autocorrelation based statistics. They were found to require more than 10 times the mean computation time of all non-autocorrelation based SQIs requiring  $43.7 \pm 0.9$  ms to be computed on a 3.1 GHz Dual-Core Intel Core i7. The reason for this increase is explained in 6.4.

#### 4.4.5 Mann-Whitney and Wilcoxon signed-rank tests

In an initial attempt to analyse the ability of the 26 SQI tests to classify PPG data samples into the three classes of quality the training subset of data samples' SQI scores and labels were used to measure the ability of a SQI to segment A and B labelled data samples from C labelled data samples (AB vs C) as well as the ability to segment A labelled data samples from B and C labelled data samples (A vs BC). This ability is estimated by both the Man-Whitney and Wilcoxon signed-rank test when performing a two-sided null hypothesis test as explained in section 2.6.4.

Figures 4.7 4.8 show the result of these tests for each 2 class segmentation tested. It must first be noted that as postulated in section 2.6.4 the segmentation scores do vary between the Man-Whitney and Wilcoxon signed-rank test that provides evidence for the interdependence of data samples however this variance in the relative order of the SQIs  $p$  value was minimal and more prevalent for larger  $p$  values.

The autocorrelation based SQIs were observed to have the highest segmentation ability for both 2 class segmentations tested where the the value of the second peak in the autocorrelation of a data sample was found to have the lowest  $p$  value in AB versus C segmentation and the lag that

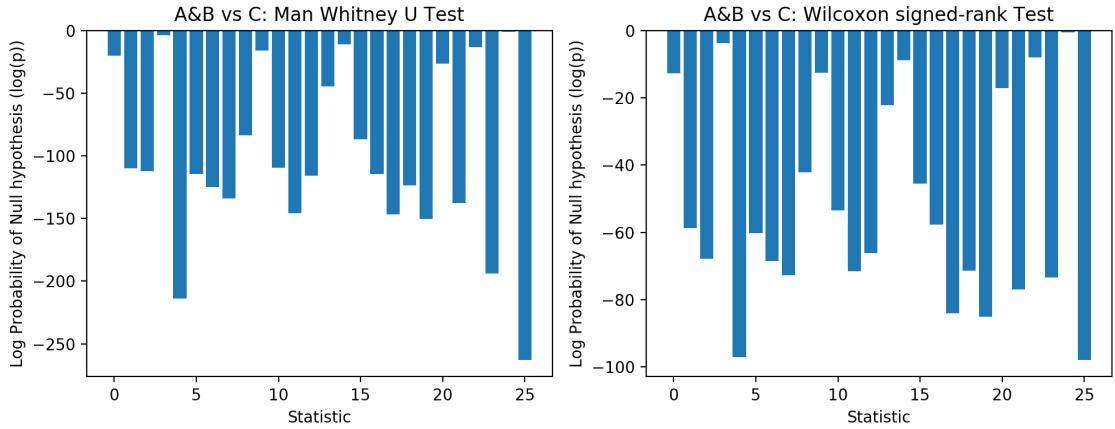


Figure 4.7: AB vs C Man-Whitney and Wilcoxon signed-rank test  $p$  values for all SQIs

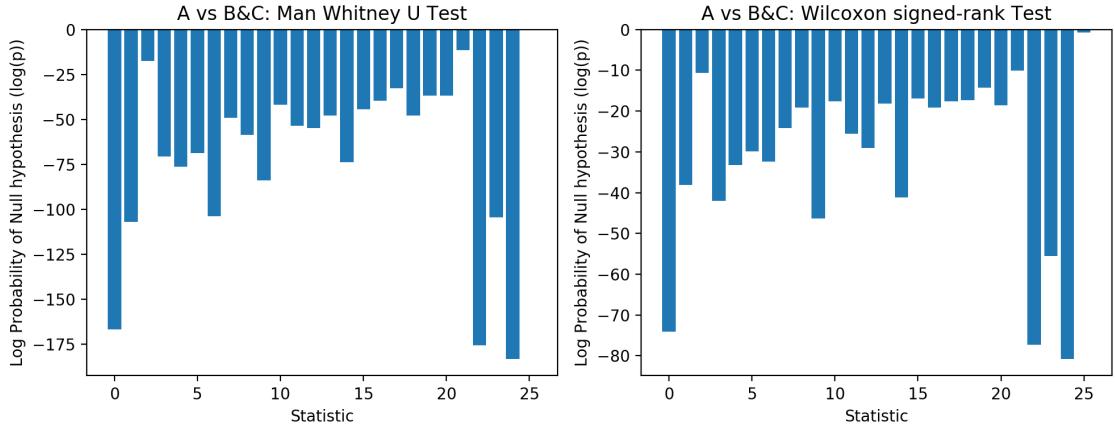


Figure 4.8: A vs BC Man-Whitney and Wilcoxon signed-rank test  $p$  values for all SQIs

corresponded to said location was found to have the lowest  $p$  value in B versus C segmentation with lag that corresponds to the first peak in autocorrelation close behind.

Discounting the autocorrelation based SQIs and providing validation of Elgendi et Al's findings in [56] Skewness was found to have the next lowest  $p$  value when segmenting A vs BC data samples, however, this was interestingly not the case when segmenting AB vs C samples, instead MSQ was found to have the next lowest  $p$  value providing validation of [70]. This analysis was missing from Elgendi et Al's work yet is of paramount importance as it effectively measures the ability of a SQI to segment acceptable signals (Classes A and B) from unacceptable signals (Class C).

#### 4.4.6 Feature importance's using a Random Forest Classifier

Although the Man-Whitney and Wilcoxon signed-rank tests provided an indication of an individual SQI's ability to segment the data samples such tests do not consider the ability of a combination of features to classify the data samples into the three different classes. Hence the next test conducted to evaluate the SQIs classification ability consisted of fitting an extra random forest classifier as explained in section 2.6.5 on to the training data such that to evaluate the relative feature importances.

Figure 4.9 shows the results of this test where the SQI numbers correspond to those outlined in Table 4.2. At first glance the test provides further evidence of the accuracy of the autocorrelation based SQIs where the value of the first peak in the autocorrelation of a data sample ranked highest in relative importance followed by the location of the second peak. The third highest ranked SQI was MSQ providing further evidence of its ability to classify signals, however, skewness, the SQI proposed by Elgendi in [56] ranked 13th. A closer inspection of the standard deviations of these mean importance's show that the results of this ranking are highly variable and when the experiment was run several times for different initial random states all places after the third highest

ranked SQI were variable.

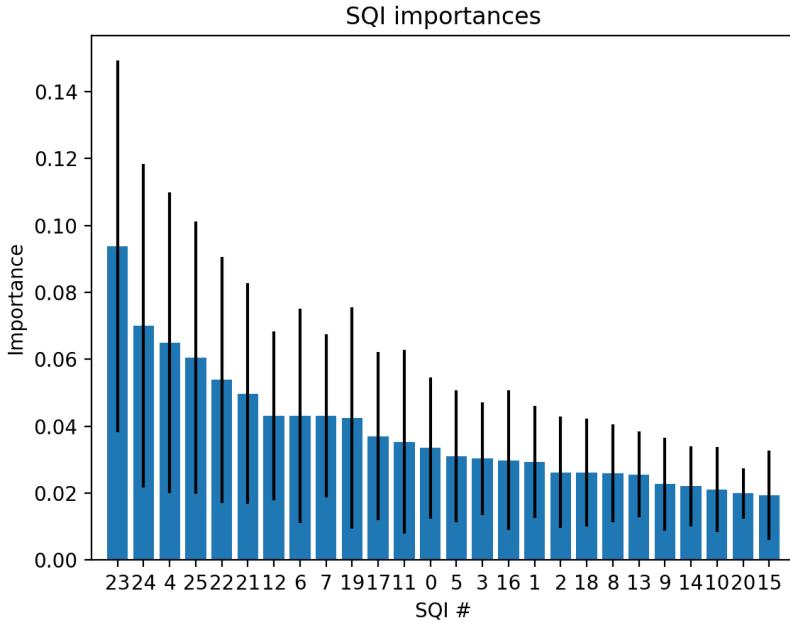


Figure 4.9: Extra trees mean relative importance of SQIs with standard deviations

#### 4.4.7 SQI selection

After conducting three tests to estimate the relative ability of the 26 reviewed SQIs, as well as a test to review their relative computational intensities, Skewness and MSQ were selected as the two features to form a multi-class linear classifier of the data samples.

The autocorrelation based SQIs were not considered as a result of their large computational requirements for a relatively small increase in segmentation ability. This decision was made in an effort to abide by the requirements outlined in section 3.1.5 in which it was stated that an emphasis on the moderation of the computation time of the algorithm was paramount to its development.

The per pulse SQIs were not considered as although several pulse SQIs ranked higher than skewness in the forest of trees feature importance test, this ranking was highly variable and skewness ranked highest in the A vs BC Man-Whitney and Wilcoxon signed-rank tests. The linear dependence of the per pulse SQIs computation times to the pulse rate of the person from which a sample will be acquired was deemed undesirable as it did not allow for an accurate profiling of the timing requirement of the algorithm and hence limited the algorithms potential real time implementation and integration into existing time sensitive systems.

The SQI feature selection was limited to two features such that to allow for a linear classification of samples such that after extracting the thresholds of these classifications the algorithm could be simplified as two linear thresholds to test for and hence significantly reduce the computational intensity of the algorithm , as well as allow its implementation in more low level environments such as on any future wearable developed as the goal of project by the Centre of Bio-inspired technology.

## 4.5 Finding Linear Classification Thresholds

As Skewness and MSQ have now been identified as the two SQIs selected for a linear 3 class classification of the data samples an evaluation on the method used to attain the thresholds for said classification was subsequently conducted. This analysis was missing from Elgendi et Al's investigation on the utility of SQIs to linearly classify data, where the threshold for classification was instead quoted empirically [56].

#### 4.5.1 Plot of training data in feature space

A plot of the training data segmented as outlined in section 4.4.1 in the new two dimensional feature space (MSQ and Skewness) selected in section 4.4.7. This was completed to provide visual intuition into the optimal method for obtaining linear thresholds for classification and to visibly confirm the effectiveness of the pre-processing method in removing unwanted variance between data samples acquired on different devices.



Figure 4.10: Feature space plot of training data

Figure 4.10 shows this plot. A legend is provided to help identify data samples within the training set that belong to the three different subsets and it can clearly be observed that there is no clear variance in the distribution of data samples from different subsets whereas there does appear to be a discernible difference in the distribution of the three classes, this provides evidence of the success of the proposed pre-processing technique in removing said inter-subset variance whilst retaining inter-class variance.

#### 4.5.2 Support Vector Machines

As explained in section 2.6.8 SVMs are a popular method for data classification and have been successfully used in the literature for multi-class classification of PPG signals using a one versus other approach to multi class classification. Once fit on training data with only two dimensions for a three class output a linear kernel SVM essentially consists of two lines within the feature space, the intersections of which coinciding with the thresholds of decision. It hence provided a systematic approach to the setting of the thresholds within the algorithm where the lines within the feature space are decided as per the explanation in section 2.6.8.

A linear SVM was fit on the training data using the MSQ and Skewness SQIs and the scikit learn python library[89]. As noted in section 4.3 as the dataset acquired had an uneven distribution of the three different classes the class weights of the fit were adjusted such that to balance this uneven distribution.

The optimal regularisation parameter  $\lambda$  was found using 10 fold cross-validation to decide upon the  $\lambda$  with the highest macro-average F1 score explained in section 2.6.7. The cross-validation

technique was preferred to a unique validation subset as a result of the relatively small size of the training set which when accompanied with the smaller distribution of classes A and C signals results in a validation set that will not accurately represent the data.

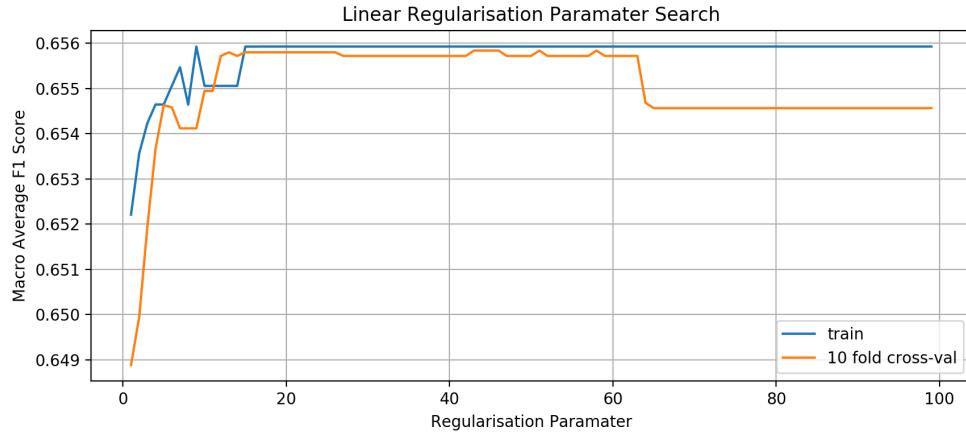


Figure 4.11: SVM regularisation parameter investigation

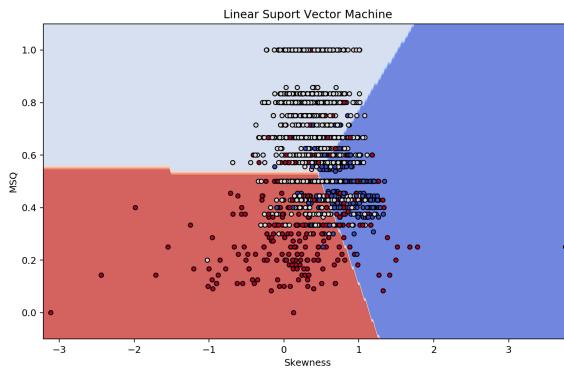


Figure 4.12: Linear SVM thresholds on training set

The results of this investigation are outlined in Figure 4.11 where the optimal  $\lambda$  was identified to be 43. The oscillation in results and the points within the curve where the cross-validation was found to provide a higher F1 score can be explained as the effects on macro-average F1 score as a result of a lack of individual class predictions. The plot of the training data segmented by the linear thresholds can be found in Figure 4.12 resulting in a 10-fold cross-validation weighted average F1 score of 65.58 %.

Although not originally planned to be conducted an investigation into other kernel types was conducted to investigate the loss in accuracy as a result of the imposed linear restrictions. Figure 4.13 shows the resulting thresholds of said SVMs when optimised for their regularisation parameter the results of which can be found in Figure A.2 of the appendix.

Although Figure 4.13 suggests that the accuracy of the model may be reduced by the linear approximation, the cross-validation F1 scores of said kernels which can be found in Table 4.3 and visually compared in Figure 4.16 which include the standard deviation of this 10 fold cross-validation show that although there is a small increase in score when using a RBF kernel the standard deviation of this score increases, this suggests that a linear model may in fact be optimal which in turn suggests an inherent linear relationship between the Skewness and MSQ indices and the quality of PPG signal segments. This relationship is postulated further in section 6.2.

#### 4.5.3 Linear Classifiers with different loss functions

Although SVMs were proven utile for linear classification, margin maximisation using a hinge loss may be limiting the models ability to fit the training data. As a result a linear classifier was investigated using 6 different losses explained in section 2.6.8.

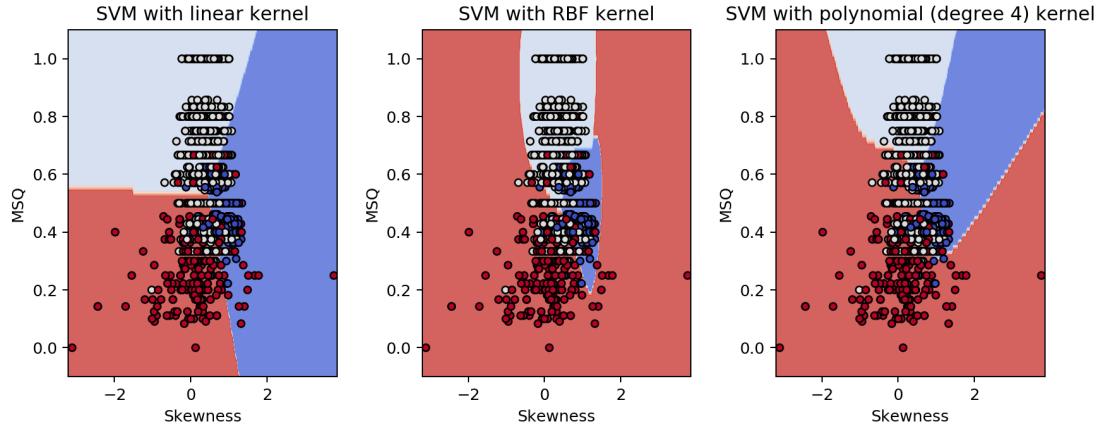


Figure 4.13: SVM kernel investigation

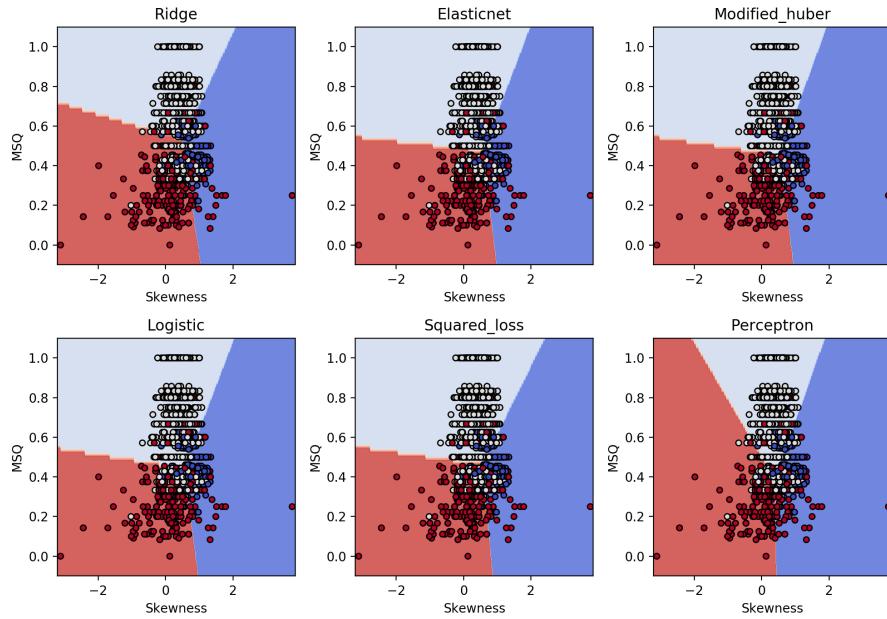


Figure 4.14: Linear classifier loss function investigation

As per section 4.5.2 where a regularisation hyper-parameter was required in each of the new models it was selected by a 1 dimensional search of the optimal value in an empirically derived range using 10-fold cross-validation, the results of these 1 dimensional searches can be found in Figure A.2 of the Appendix.

Reassuringly the general shape of the thresholds as observed from Figure 4.14 do not appear to change dramatically providing evidence of the robustness of the classification where instead minor changes to the gradient and offsets of the linear thresholds are observed.

The training and 10 fold cross-validation scores for the classifiers in 4.14 can be found in Table 4.3 and visually compared in Figure 4.16. It can be noted from Figure 4.16 that the standard deviation of the cross-validation F1 score for the perception classifier function was substantially larger then that of the other models, furthermore, its cross-validation F1 score is also lower than the other models. A closer inspection of the perception classifier's outputted thresholds in Figure 4.14 suggests that the perception model is overfitting the training data, explaining the reduction in cross-validation score and increase in standard deviation and justifying the need for an optimised regularisation parameter when setting linear thresholds for classification.

#### 4.5.4 Shallow Decision Tree

The final linear classifier evaluated for its use in deciding upon the thresholds of classification was a shallow decision tree as explained in section 2.6.8. The thresholds selected by the decision tree when restricted to a depth of two essentially result in four, one dimensional thresholds for classification. As the annotation method classifies the data into one of three classes corresponding to three segments of the feature-space. Although four thresholds must be checked for such classification as oppose to the three required by the previous linear classifications this type of one dimensional classification is less computationally intensive than that of the linear models tested so far as it does not require a weighted sum of the computed scores to be calculated instead only requiring to check if either Skewness or MSQ are larger than or smaller than a specific value at most four times.

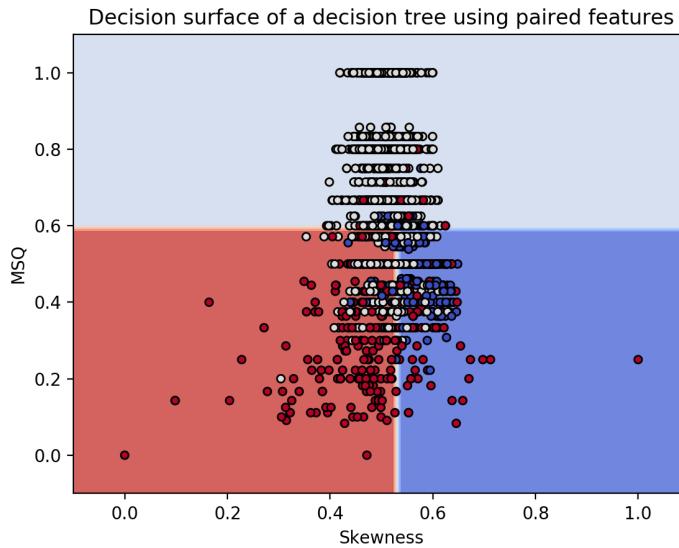


Figure 4.15: Fitted decision tree classification thresholds

The thresholds of the fitted decision tree can be seen in 4.15. The scores of this tree can be found in Table 4.3 and visually compared to the other classifiers in Figure 4.16 where it is clear that there is a substantial loss in accuracy as a result of this simplification.

#### 4.5.5 Selected linear threshold

Table 4.3 outlines the scores of all the linear classifiers tested. The highest mean cross-validation F1 Score corresponded to a linear classifier with an Elastic Net penalty for regularisation this is likely as a result of the fact that Skewness and MSQ are inherently both attempting to measure a PPG signal samples' quality and will therefore be correlated and as mentioned in section 2.6.8 Elastic Net loss functions are useful for such situations. An inspection of the standard deviation of this score however, reveals it be comparatively higher than the other models. In particular the linear classifier with least square regression which achieved a comparatively high Cross Validation F1 score with a slightly smaller standard deviation of the result. All of the models were found to suffer from the effect of outliers and those that were designed to be tolerant towards outliers were found to perform better. The square loss function penalizes outliers heavily [103], explaining its lower standard deviation to variations in cross-validation data and its comparatively higher mean cross-validation score. It was hence the selected model for providing the linear thresholds for the algorithm.

Figure 4.19 shows the process flow of the final algorithm for the three class classifier from raw signal input to classification output including the thresholds extracted from the linear classifier with least square regression.

Classifier	Train Score (%)			Validation Score (%)		
	Precision	Recall	F1	Precision	Recall	F1
SVM linear	71.37	63.92	65.59	$71.34 \pm 3.22$	$63.94 \pm 2.40$	$65.58 \pm 2.66$
SVM RBF	72.78	66.19	67.62	$72.48 \pm 3.52$	$66.02 \pm 3.18$	$67.27 \pm 3.17$
SVM polynomial	70.32	61.11	62.37	$69.97 \pm 3.33$	$60.92 \pm 2.25$	$61.94 \pm 2.81$
Ridge	70.80	63.89	65.78	$70.78 \pm 2.97$	$63.99 \pm 2.28$	$65.71 \pm 2.47$
Elasticnet	68.58	66.94	67.02	$69.73 \pm 4.54$	$67.08 \pm 4.14$	$67.46 \pm 4.02$
Mod Huber	68.04	66.99	66.54	$68.14 \pm 3.30$	$66.98 \pm 3.58$	$66.69 \pm 3.13$
Logistic	66.47	67.11	66.16	$68.72 \pm 3.51$	$67.02 \pm 3.77$	$66.94 \pm 3.16$
Squared	67.29	67.01	66.40	$68.19 \pm 3.27$	$67.61 \pm 3.73$	$67.34 \pm 3.30$
Perceptron	66.86	68.27	64.87	$56.68 \pm 8.69$	$65.23 \pm 10.58$	$53.81 \pm 11.84$
Decision Tree	66.86	68.27	64.87	$56.68 \pm 3.18$	$65.23 \pm 2.14$	$53.81 \pm 2.47$

Table 4.3: full table of results for three-class linear classification investigation

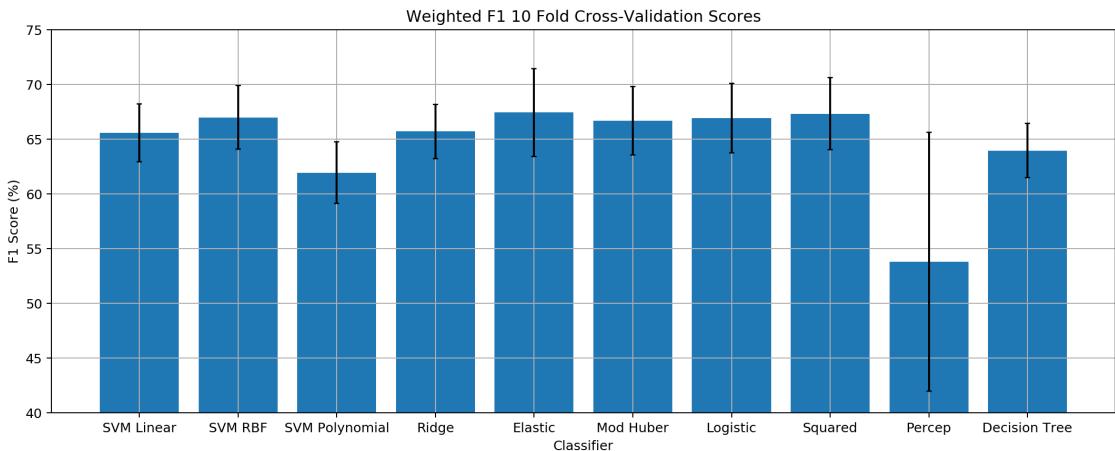


Figure 4.16: 10-fold cross-validation F1 scores of the classifiers with standard deviation for three-class classifier

#### 4.5.6 Two-class classifier

To increase the utility of the graphic user interface for PPG signal quality classification and to further illustrate the use of the skewness and MSQ scores to segment samples of varying quality, Samples of Class A and B were grouped together as a group of samples deemed to contain PPG pulse waveforms acceptable for analysis as appose to the samples of Class C that contain pulse waveforms unsuitable for analysis. This was done such that a new two-class classifier was fit to the training data. This classifier was developed and implemented such that to increase both accuracy and computation time for situations in which the user does not require the identification of class A signals and instead only requires unacceptable signal samples to be identified an example is simply when identifying a viable ppg signal during initial probe placement or calibration.

To find the optimal linear classification thresholds for this new two-class classifier the regularisation parameter of all the linear classifiers tested previously were once more optimised the plot of which can once more be found in Figure A.3 of the appendix.

From Table 4.4 and Figure 4.17 it is interesting to note that the relative order in performance of the classifiers remains the same for the two class classifiers as it was for the three class classifiers. This provides evidence for the robustness in the selection of the Squared Loss function linear classifier for threshold selection for the given scores as the scores used to segment the data have remained the same and hence the arguments outlined in section 4.5.5 for why the Squared Loss classifier suited the feature-space extends to a different arrangement of the classes to be segmented within the feature-space.

Figure 4.18 shows the threshold extracted from the linear classifier to classify the training data and Figure 4.19 shows the process flow of the final algorithm for the two class classifier from raw signal input to classification output including the thresholds extracted from the linear classifier with least square regression.

Classifier	Train Score (%)			Validation Score (%)		
	Precision	Recall	F1	Precision	Recall	F1
SVM linear	75.26	76.20	75.71	$75.52 \pm 5.15$	$76.66 \pm 4.81$	$75.98 \pm 4.87$
Ridge	75.82	67.63	68.85	$75.94 \pm 4.50$	$67.71 \pm 2.68$	$68.82 \pm 2.98$
Elastic Net	75.51	76.00	75.75	$75.47 \pm 4.99$	$76.16 \pm 4.53$	$75.74 \pm 4.66$
Mod Huber	76.77	71.29	73.17	$76.13 \pm 3.93$	$72.19 \pm 3.84$	$73.68 \pm 3.98$
Logistic	76.74	70.82	72.74	$76.01 \pm 4.03$	$71.99 \pm 4.01$	$73.48 \pm 4.15$
Squared	77.54	70.45	72.42	$77.70 \pm 4.76$	$72.91 \pm 4.37$	$74.63 \pm 4.62$
Perceptron	76.70	78.58	77.58	$70.10 \pm 8.18$	$73.27 \pm 23.02$	$67.90 \pm 18.17$

Table 4.4: Full table of results for two-class linear classification investigation

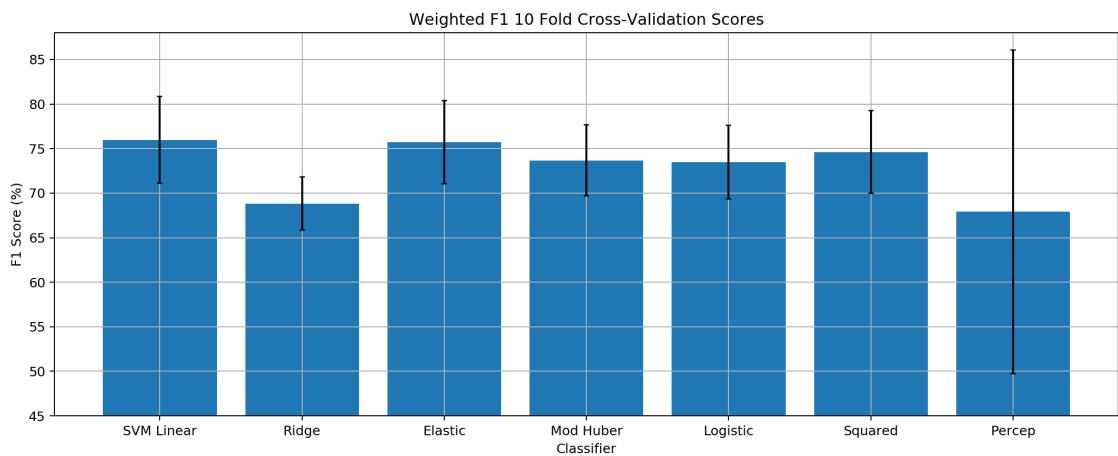


Figure 4.17: 10-fold cross-validation F1 scores of the classifiers with standard deviation for two-class classifier

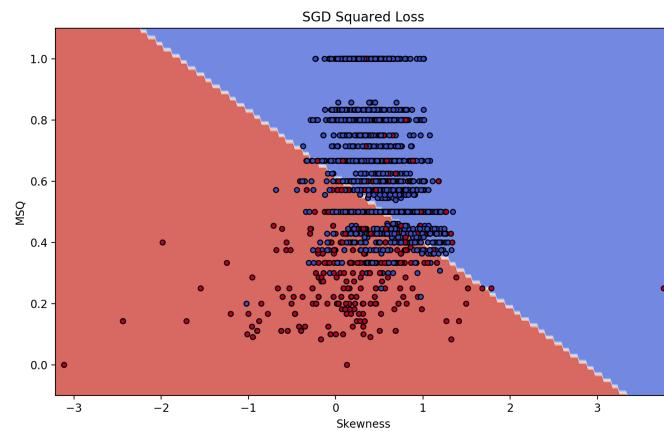


Figure 4.18: Squared loss two-class classifier segmentation

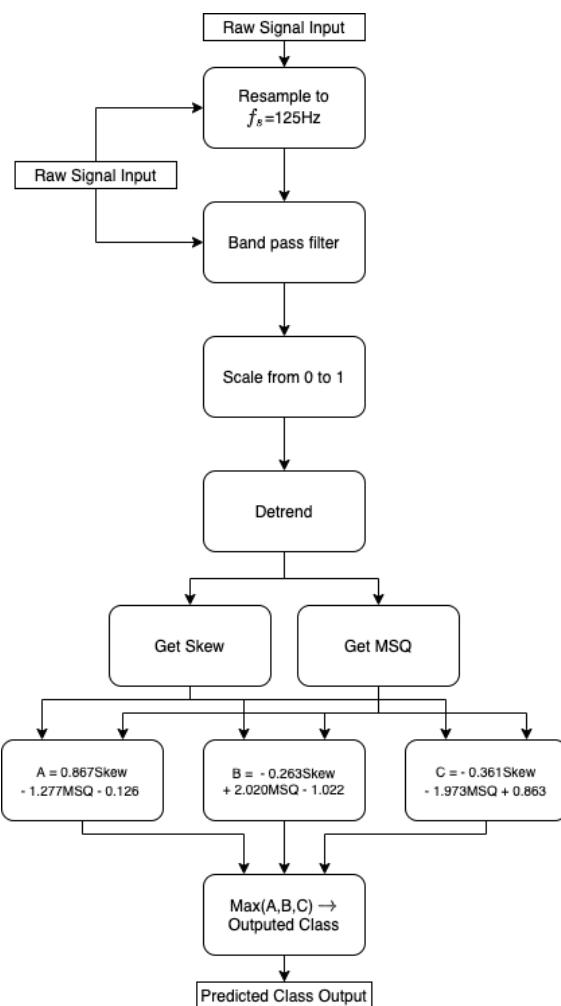


Figure 4.19: Diagram illustrating three-class classifier algorithm

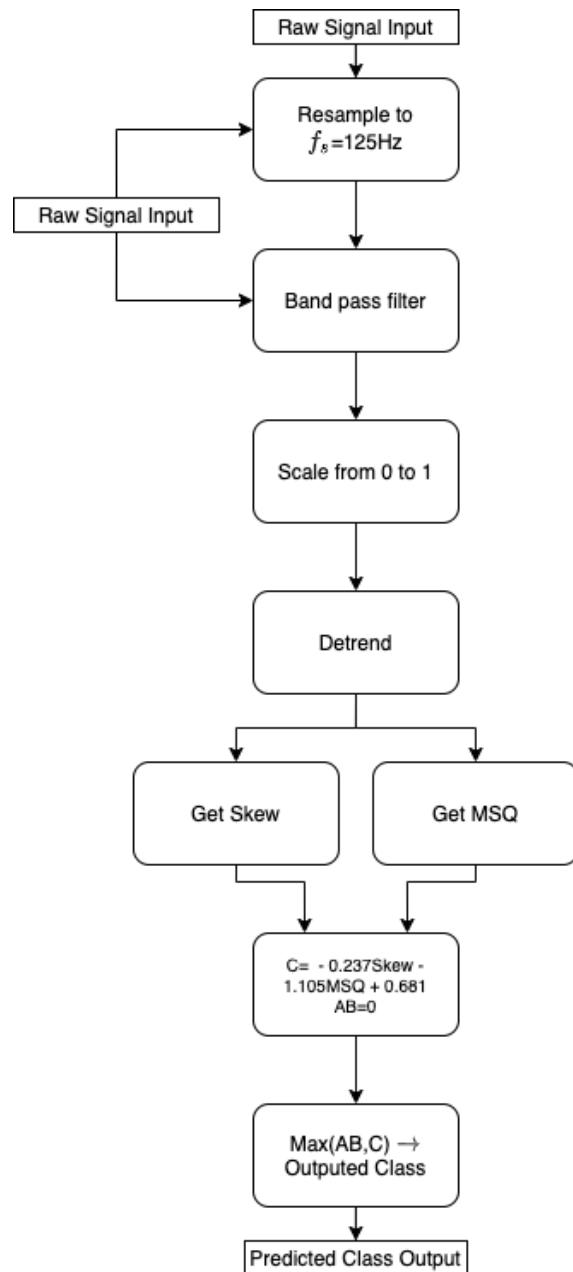


Figure 4.20: Diagram illustrating two-class classifier algorithm

# Chapter 5

## Implementation

### 5.1 Algorithm Implementation into Graphic User Interface

After the development of the algorithm to fulfil the requirements outlined in section 3.2 and in an attempt to showcase its use, a graphic user interface (GUI) was developed capable of importing a signal to be classified, or recording PPG signals from a AFE4900EVM and classifying said signals in real time with the option to analyse the signal afterwards.

This GUI was implemented in python to directly interface with the AFE4900EVM python driver currently being developed by Stefan Karolcik at the Centre for Bio-Inspired Technology. The PyQt library [104], a Python binding for the Qt cross-platform C++ framework, was used to develop the varying widgets of the GUI as the library although simplistic allowed for all desired features to be implemented and was sufficiently computationally inexpensive. Data acquisition, analysis and display are all conducted in separate threads where queues were implemented to allow for the distribution of data from one thread to another. A Multi-thread approach was utilised to improve the computational intensity of the graphic user interface by separating computations that could be run in parallel. This allowed for the real-time, updating display of the raw PPG signals acquired by the AFE4900 and will allow for immediate visual feedback of any found PPG signal, significantly improving a trained individual's ability to place any ppg probes correctly.

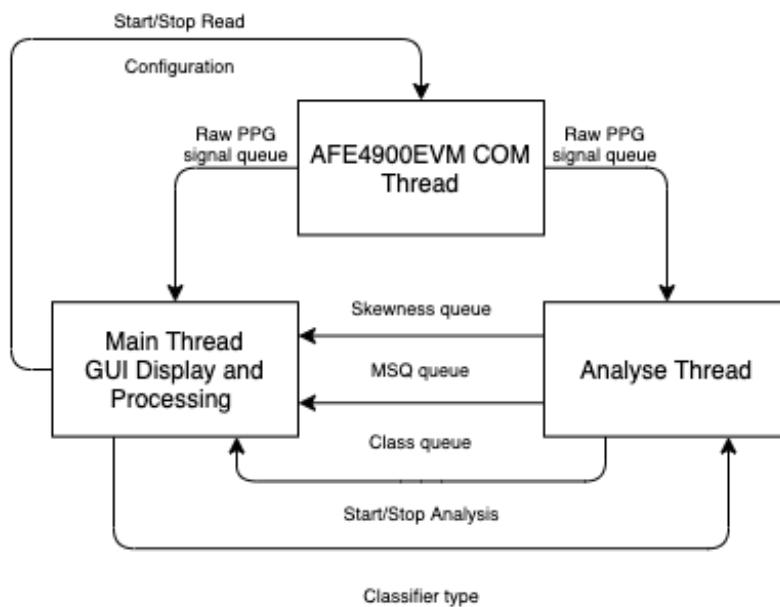


Figure 5.1: Diagram of thread communication

Figure 5.1 shows how these queues allow for the distribution of data between the three defined threads, data acquired by a thread dedicated to interfacing with the AFE4900EVM and receiving PPG signal data through the com port is placed in two separate queues to be received by the display and analyse threads respectively. The analyse thread collects three second samples of raw

ppg signals and computes the MSQ, skewness and classification of each sample, placing them into three separate queues to be received by the display thread which displays the real-time raw ppg signal as well as the MSQ and skewness scores as well as the predicted class. The GUI thread which is the main thread of the program is also responsible for controlling the other two threads and providing their necessary parameters for operation, for example the classifier type used in the analysis thread is selected by a user using the GUI and this configuration is subsequently passed to the analyse thread. The configuration of the AFE4900EVM is also selected by the user using the GUI and passed to the COM thread.

Figure 5.4 shows the flow of the graphic user interface developed where rounded boxes indicate different widgets within the GUI and square boxes indicate inputs or outputs. It must also be noted that it is possible from any screen to revert back to the File selection or Sensor selection widget using the menu at the top of the GUI. Once a signal is classified it is possible to export the raw signal, filtered signal and annotations to a csv file. It is possible to export any available data at any point after acquisition regardless of whether it has been filtered or classified.

For both real-time classification and classification from an imported signal it is possible to select either the three-class output algorithm or the two-class output algorithm. This was implemented to easily suit the requirements of the user where if it is desirable to identify class A signals for more complex feature analysis using the discrotic features of a patients PPG signal the user may use the default algorithm. On the contrary if it is desired to simply identify unacceptable signal segments corrupted by unwanted artefacts with a higher accuracy then the two-class classifier can be selected instead.

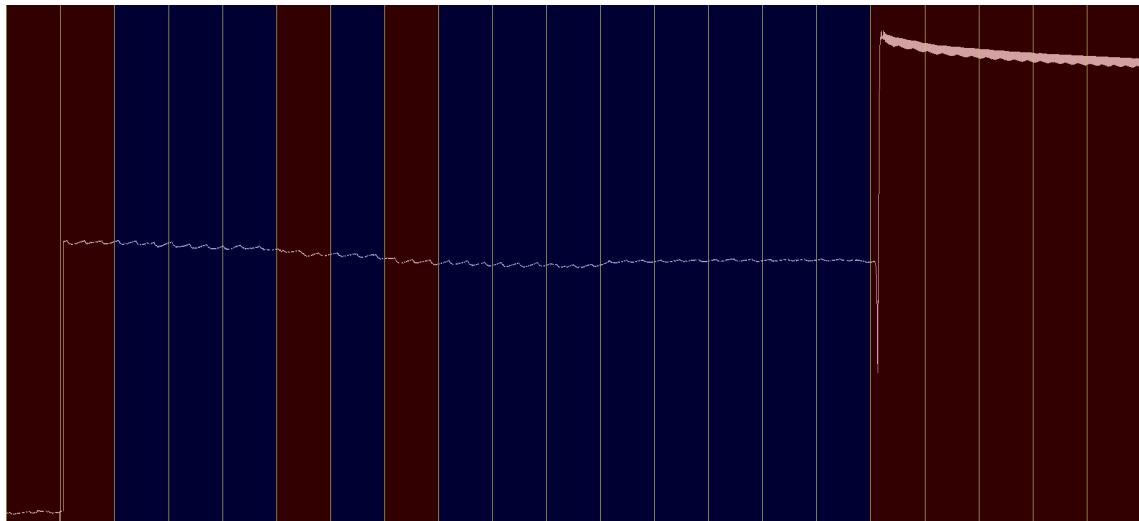


Figure 5.2: Annotation of signal acquired and analysed in real-time

Figure 5.2 shows an example of a segment of data acquired by the AFE4900EVM that was classified in real-time using the algorithm illustrated in Figure 4.20. Segments annotated with a blue background correspond to segments that were classified as class B and segments annotated with a red background correspond to segments that were classified as class C. It is clear from Figure 5.2 that the algorithm successfully identifies segments of the acquired data where unwanted artefacts introduced as a result of probe removal and calibration are present in real-time.

Figure 5.3 shows an example of a segment of data imported to the GUI classified using the algorithm illustrated in Figure 4.19. In this Figure one can observe segments annotated with a green background corresponding to segments that were classified as class A. Figure 5.3 shows that the algorithm successfully identifies segments of the acquired data where both the systolic and diastolic peaks are present within the data as class A signals.

## 5.2 Tools for Further Algorithm Development

Chapter 4 outlined the development of a universal framework for real-time classification of three second samples of a PPG signal from any PPG acquisition device. Although classification algorithms have been developed with universal thresholds deemed to provide sufficient accuracy, tools

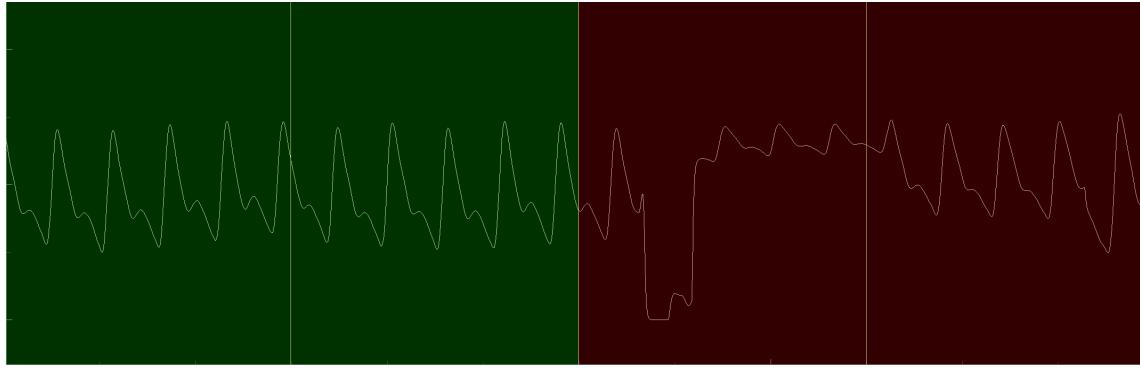


Figure 5.3: Annotation of signal imported and analysed

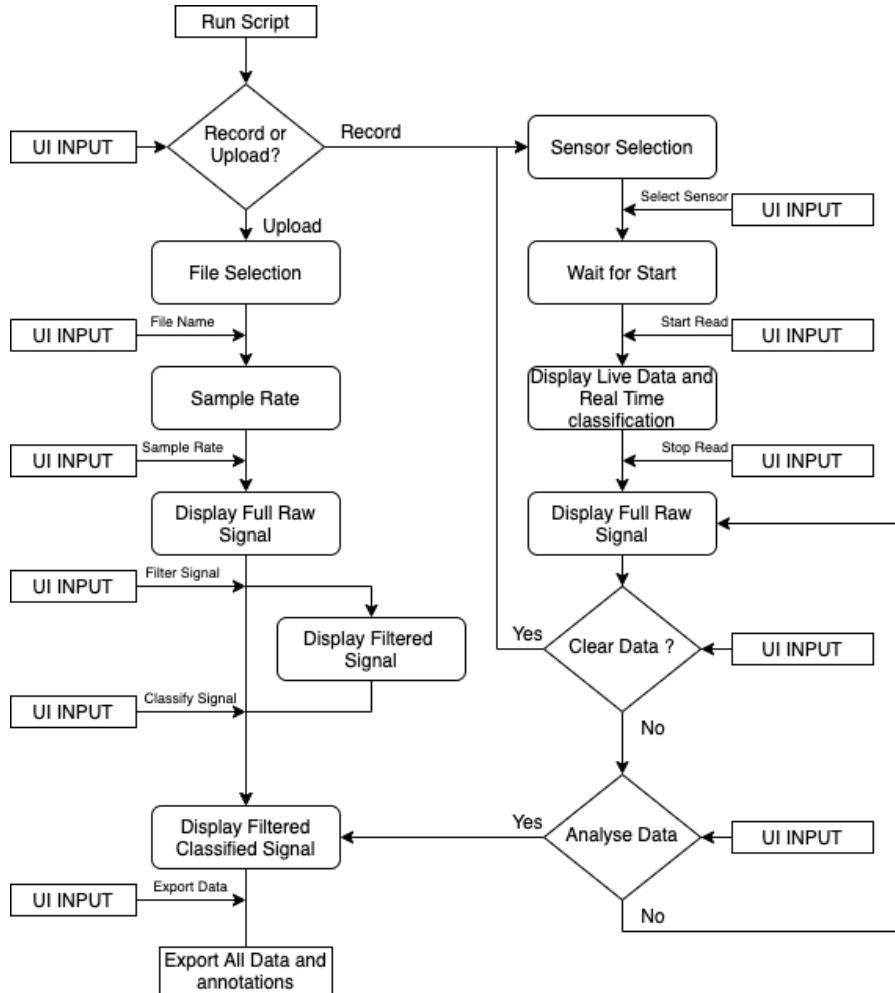


Figure 5.4: Flow of GUI created

to allow the simple optimisation of these thresholds for a given acquisition device provided a dataset of existing signal samples.

The first of which is a GUI for the manual annotation of said signal samples into one of three classes. A screenshot showcasing the GUI can be found in Figure 5.5 which shows the raw or filtered PPG signal sample depending on the toggle button "Filter" and allows the user to select its class using one of three class buttons, skip to the next signal and go back to the previous sample. The GUI also allows an annotator to save the current annotations such that the application can be quit and annotation can be resumed where it was left off another time. Finally the user can export these annotations with a user specified file name into the root directory.

The second tool was created in the form of a guided jupyter notebook that when provided with an annotation file and the signal samples attained using a specific device or specific set of

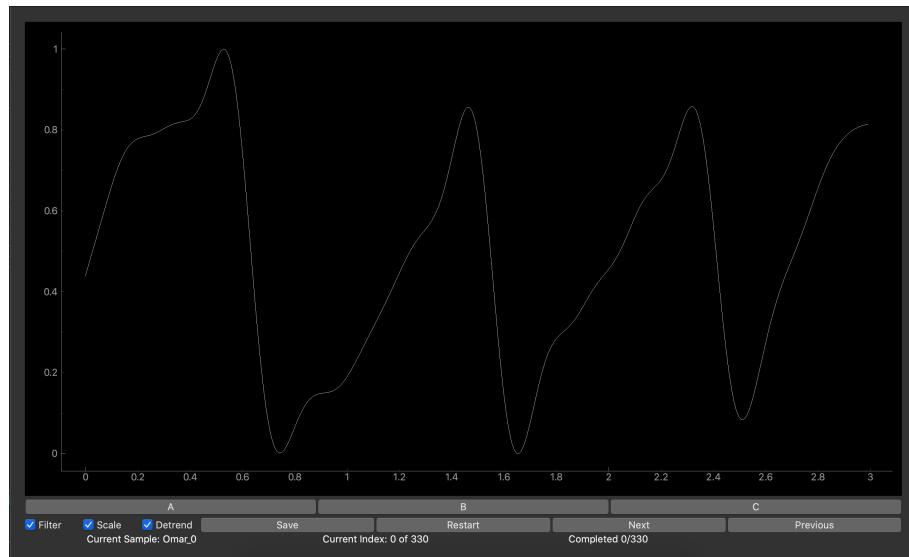


Figure 5.5: GUI for dataset annotation

devices, uses the framework outlined in chapter 4 to provide thresholds for the linear classification of samples optimised for said device(s). These thresholds can be passed to the Graphic User Interface outlined in section 5.1 when classifying an imported signal.

# Chapter 6

## Testing and Results

### 6.1 Testing Accuracy of Implemented Algorithms

#### 6.1.1 Three-class classifier

To test the accuracy of the implemented algorithms a test set composed of data not used in any of the analysis and development stages was used. Details of the composition of this test set can be found in section 4.4.1.

	Precision	Recall	F1	# Samples
A	57.21	48.97	52.77	243
B	84.53	82.65	83.58	853
C	49.43	62.09	55.04	211
Accuracy			73.07	1307
Macro average	63.73	64.57	63.80	1307
Weighted average	73.79	73.07	73.25	1307

Table 6.1: Testing classification report of three-class classifier

Figure 6.1 shows the classification results of the proposed three-class classification algorithm from Figure 4.19. It is clear from the results that the algorithm excels in the classification of Class B signals and the macro average F1 score used to develop and decide upon thresholds for classification in section 4.5 suffers heavily as a result of the F1 scores for classifying class A and class C signals. An inspection of the number of samples within the test shows a similar distribution to that of the training set where class A and class C signals are unevenly represented. The weighted average f1 score which considers this uneven representation and considered to better represent practical performance [57] was found to be 73.35%. Although the macro average f1 score was selected in development to equally penalise classification errors from all three-classes, the weighted average f1 score is more indicative of the accuracy of the algorithm given the general distribution of PPG signal quality data.

A confusion matrix which further outlines the performance of the three-class algorithm is shown in Figure 6.1. It is interesting to note that the majority of miss labelled signals of class A were classified as class B. This bodes well for the overall accuracy of the classifier as said signals would still be considered viable in any application of the algorithm. On the contrary however signals of class B that were miss labeled were almost equally labeled as class C or A. An ideal classifier would have a bias towards accurately classifying signals of class A and hence allowing closely thresholded signals to be of class C such that prioritise the correct identification of high quality signals. It must be noted that this is achievable when fitting linear classifiers as the weights of the classes can be adjusted accordingly, however, as previously mentioned the weights of the classes were adjusted to balance the three-classes as a result of their uneven distribution within the dataset.

#### 6.1.2 Two-class classifier

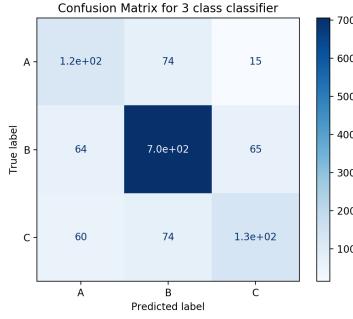


Figure 6.1: Confusion matrix of test set for three-class classifier

	Precision	Recall	F1	# Samples
<b>AB</b>	91.07	89.44	90.25	1061
<b>C</b>	57.74	62.20	59.88	246
<b>accuracy</b>			84.32	1307
<b>macro avg</b>	74.41	75.82	75.07	1307
<b>weighted avg</b>	84.80	84.32	84.54	1307

Table 6.2: Testing Classification report of two-class classifier

The classification of the proposed two-class classifier on the testing set can be found in Table 6.2. The effect of the distribution of the training data on the bias of the classifier is clear once more from the uneven distribution in precision and recall score for the two-classes however the precision, recall and subsequently the F1 score of the samples of class C was found to increase slightly. The majority of the increase in the macro average F1 score of the classifier when compared to the three-class classifier occurs as a result of the increase in AB class classification. This increase is explained by the observation made in section 6.1.1 with reference to Figure 6.1, where the majority of mislabeled A class signals that were classified as class B by the three-class classifier are now correctly classified as class AB. The weighted average f1 score of the classifier which better represents its performance in practical application [57] was found to be 84.54 %.

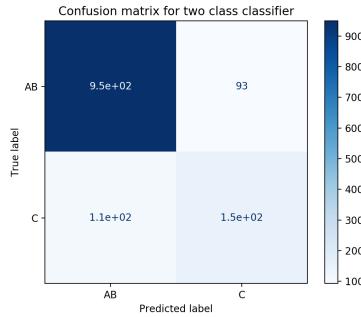


Figure 6.2: Confusion matrix of test set for two-class classifier

Figure 6.2 which shows the confusion matrix of the three-class classifier once more highlights the problem outlined in section 6.1.1 where there is a considerable number of signals labeled as Class C that were predicted to be of class AB.

## 6.2 Feature Space Distribution analysis

One can further understand the results of the confusion matrices in Figures 6.1 and 6.2 by viewing the distribution of the three-classes within the 2 dimensional feature space used to classify the samples as per Figure 4.10. From Figure 4.10 it is clear that signals of class A had MSQ scores between 0.4 and 0.6 however high skewness scores as per [56]. As a result it is clear from Figures 4.12, 4.13, 4.14 and 4.15 that linear classifiers trained on said distribution classified A class signals with a low MSQ however a high skewness as class A. Class C signals were also classified as having

a low msq but a low skewness. It hence would only require 1 dimension of the feature-space (skewness) to increase such that to cross over the threshold and change a signals classification from C to A.

The distribution of the MSQ scores of samples of label A around 0.5 can be explained as a drawback in the design of the MSQ score where the Scipy peak detection algorithm [73] identifies both systolic and diastolic peaks whereas Billauers' algorithm was only identifying systolic peaks and hence this mismatch of approximately half the peaks not identified by Billauers algorithm for A class samples results in MSQ scores of approximately 0.5. An example of this mismatch can be found in Figure 6.3.

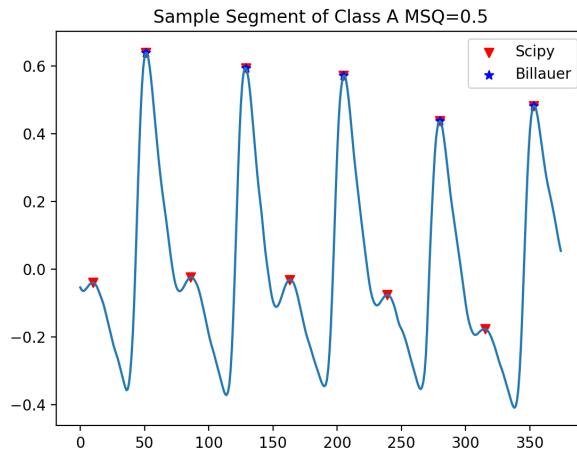


Figure 6.3: MSQ class A example

### 6.3 Testing Computation Time of Implemented Algorithms

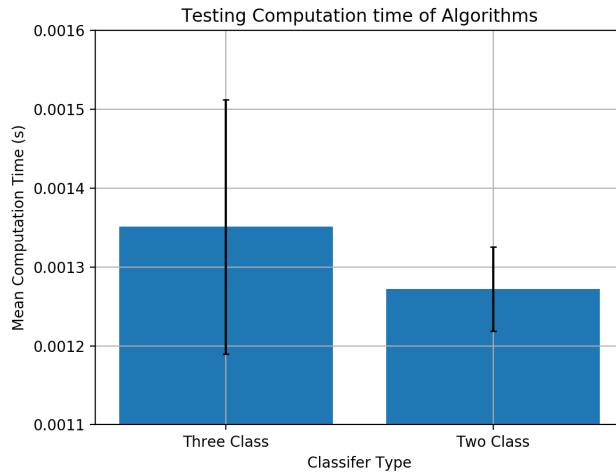


Figure 6.4: Computation time of algorithms

To assess the developed algorithms speed, the time taken for filtration, calculating the MSQ and skewness score and finally classification was measured once more using the "timeit" python package [101]. This was conducted on 40 random samples of data from the training set such that to incorporate raw signals of varying sample rate and hence varying raw length for the same 3 second segment. For each sample a further 100 tests were conducted providing a mean time of computation as well as a measure of the variance of this value per sample. The mean times of computation per sample were averaged together to provide a mean computation time for the algorithm in practical

application and the variance of the samples were averaged and square rooted to provide an estimate of the standard deviation of this mean computation time.

Figure 6.4 shows the results of these tests and it is clear that both algorithms require minimal computation time when implemented in python and tested on a 3.1 GHz Dual-Core Intel Core i7 processor where both classifiers required less than 2ms to filter, compute scores and classify the data-samples in real time. It must be noted that this absolute value of computation time is processor dependent and will change subject to the device and language the algorithm is being run on but nonetheless classification was found to have no effect on the speed of the graphic user interface when reading and displaying data from the AFE4900EVM both of which are computationally expensive but can therefore be prioritised.

## 6.4 Computational Complexity of Signal Quality Indices reviewed

SQIs 0-5 from table 4.2 are of computational complexity  $O(n)$  as they require only a single pass through a signal sample for their computation. This explains their relatively low computational times in Figure 4.6. Relative power is of computational complexity  $O(n \log(M))$  as it requires Welches method to perform the windowed FFT of a signal of  $M$  windows [74].

All per pulse variations of SQIS 0-5 are of complexity  $O(M \frac{N}{M}) = O(N)$  however, such analysis does not account for the overhead required to call functions several times. This explains the small increase in mean computation time in Figure 4.6. For a high heart-rate which will require the functions to be called much more times this can account for substantial increase in computation time, further validating their omission in section 4.4.2. Similarly per pulse variations of relative power are of computational complexity  $O(n \log(M))$  when not accounting for the overhead required to call functions several times.

Auto-correlation based SQIs were found to be excellent SQIs for the separation of training examples into the three annotated classes in all of the three tests conducted in sections 4.4.5 and 2.6.5, however, in section 4.4.4 the auto-correlation based SQIs were found to require considerably longer computation times relative to the other SQIs reviewed and were hence not selected in section 4.4.2 when developing a linear algorithm.

The brute force traditional method of calculation of a signals autocorrelation which was implemented for the purpose of this study is of complexity  $\alpha n^2$ . Finding the coordinates of the first two peaks is then of complexity  $\beta n$  such that the total complexity of finding the four auto-correlation based features is  $\alpha n^2 + \beta n = O(n^2)$ . As a result of the large sample length  $n=375$  this increase in the power of complexity results explains the jump in computation time required to compute the auto-correlation based SQIs apparent in Figure 4.6.

# Chapter 7

# Evaluation

## 7.1 Context

Before evaluating the proceedings of the project it must be noted that as a result of the COVID-19 pandemic the goals originally set out to be completed as part of this project were changed significantly. This is because the shutdown of all unnecessary laboratories at Imperial College and the shift to remote working resulted in the inability to conduct the project work in parallel to the development of sensors by the Centre for Bio Inspired Technology.

Originally the project goal was to further develop the acquisition platform to be used in the study in Vietnam and use it to complete small healthy patient studies to evaluate methods proposed in the literature to estimate blood pressure and hematocrit. The results of these healthy patient studies were then to be used such that to refine the acquisition platform and outline its potential use in the larger study in Vietnam.

As a result of the shift to remote learning and the inability for rapid sensor development the focus of the project was shifted towards the development of a signal quality classification tool that would both aid in PPG probe development by the centre of bio inspired technology but also help guide nurses and patients on the optimal placement of the probes in future studies as well as help to omit unusable data from analysis.

## 7.2 Meeting the set requirements

With reference to the requirements set in section 3.1 and the the framework used to develop the proposed classification algorithms in chapter 4, one can conduct an evaluation of the extent of which the requirements were met.

### 7.2.1 Meeting the algorithm requirements

The first of the requirements set out for SQI review and algorithm development in section 3.1.2 was that the dataset used for such analysis was acquired from a range of different devices at different sampling rates such that to allow for one to conclude that the findings of the analysis would extend to any PPG device. Efforts outlined in section 4.1 were made such that to meet this requirement using datasets such as the BIDMC dataset which was assumed to contain signals originally acquired using different PPG signal devices as a result of the range of different shaped PPG signals present within the dataset devices, however, the details of how many different devices used within the dataset was not found. Nevertheless at least four different PPG acquisition devices with four different sampling rates were used to form the dataset and this was deemed sufficient.

The second requirement set in section 3.1.1 was that any SQIs reviewed or algorithms developed must require no input that cannot be extracted from a single three second PPG signal sample. This requirement was addressed in section 4.4.2 and hence was met during the development of the algorithm. Section 4.4.2 also mentions that all of the SQIs were provided with PPG signal samples limited to three seconds such that to also meet the requirement outlined in section 3.1.5.

The third requirement for SQI review and algorithm development outlined in section 3.1.4 was that the utility of the findings of the study must be optimised by allowing for a wide range of their potential applications. The SQIs were reviewed for their ability to classify PPG signal

segments into one of three classes and algorithms for both two-class and three-class classification were proposed. These algorithms do not require initial calibration and hence also have the potential to identify initial incorrect ppg sensor placements.

The final requirement for SQI review and algorithm development was outlined in section 3.1.5 and required that SQIs must be reviewed for their computational intensities and a emphasis on regulating this computational intensity must be conducted in algorithm development. Tests were conducted in section 4.4.4 to measure the relative computational intensities of the different SQIs and when selecting SQIs for algorithm development in section 4.5 an emphasis was made on penalising potential SQIs for their computational intensity. The developed algorithms were then tested for their speed in section 6.3 and were found to be sufficiently computationally inexpensive.

### 7.2.2 Meeting the Graphic User Interface requirements

As outlined in section 5.1 efforts were made such that to meet the compatibility, ease of use, utility and export capability requirements outlined in sections 3.2.2 3.2.3 3.2.4. With the exception of the ease of use requirement these requirements were objective and hence can be claimed to be completely successfully met as the functionality of the GUI required to meet these requirements were successfully implemented. The ease of use requirement however is subjective and was intended to be tested by allowing volunteers to use the GUI and subsequently review its use in an anonymous questionnaire, such that the results of said questionnaire would provide one with an indication of the GUIs ease of use. Due to the shift to remote learning/working as a result of the COVID-19 pandemic this was not possible and hence no evaluation of the ease of use can be conducted.

## 7.3 Comparison of Findings to the Literature

When comparing the findings of this study to the literature this comparison was limited to the approaches outlined in section 2.5.3 as these studies met some if not all of the requirements outlined in section 3.1 and hence had similar imposed limitations and intended applications.

When conducting an SQI review of the 26 SQIs implemented on a dataset of PPG signals acquired from a range of devices, the results of section 4.4.5 when measuring the separation ability of different SQIs to isolate class A signals validated the findings of Elgendi et al in [56] which proposed skewness for this purpose. However, the limitation of this study identified in section 2.5.3 for its omission to measure the separation ability of different SQIs to isolate class A and B signals grouped together from class C, was validated as a drawback of skewness' application as a SQI. MSQ, which was disregarded in [56] was instead found to be the optimal SQI from those reviewed in [56] for such separation.

The results of Pradhan et al's findings in [57] [75] were largely validated in section 2.6.5 where autocorrelation based SQIs and per pulse based SQIs were found to be important for random forest classification. However, the variance of this relative importance was found to vary significantly between decision trees confirming the limitation identified in section 2.5.3 that Pradhan et al's findings may vary when subject to a more general dataset.

By conducting a review of the computational intensities of the PPG SQIs implemented in section 4.4.4 it was found that the auto-correlation based SQIs featured in the algorithms proposed by both Pradhan et al and Vadrevu et al in [75] and [58] respectively had high computational intensities as a result of their higher computational complexities. The omission of such an analysis in [75] and [58] was identified as a limitation in section 2.5.3 and was hence validated showing that further optimisation of an algorithms accuracy with respect to its computational requirement should be conducted.

Comparing the performance of the proposed algorithms outlined in section 4.5 and tested in section 6.1 to the performance of the algorithms from the literature outlined in 2.5.3 is difficult as different metrics were used to test the algorithms in question and each of the algorithms actually perform different classification depending on the annotation method and the classes to be distinguished, furthermore, these algorithms were developed and tested using different datasets and different methods of annotation. For example Elgendi et als optimal classifier in [56] for a sample length of three seconds was found to have an F1 score of approximately 86%, however, this is for a two-class classifier used to exclusively classify signals of excellent or acceptable quality and the evidence that a similar three-class classifier will suffer a drastic fall in performance was outlined above, the classifier was also only tested on a locally acquired dataset using only one PPG

signal acquisition device and the F1 score of the classifier is likely to reduce further for a more general dataset. The algorithms proposed in [75] and [58] attempt to classify signals as either acceptable or unfit, similar to two-class algorithm proposed in section 4.5.6. When comparing the accuracy of the two-class algorithm to that quoted by Pradhan et al for their optimal classifier (74.5%) which takes in as input a 10 second PPG signal sample, the two-class algorithm proposed in section 4.5.6 had a higher test accuracy of 84.32%. The two-class algorithm is substantially less complex than that proposed in [75] and takes in a PPG signal sample length of only three seconds. This provides evidence that the simplifications and limitations imposed in section 3.1 have not limited its performance, however, the F1 score quoted in [58] for Vadrevu et al's proposed four step hierarchical two-class classifier ( $\approx 98\%$ ) suggests that improvements to the framework used for algorithm development can be made by softening the simplifications and limitations imposed in section 3.1 as well as increasing the size of the dataset used for training.

## 7.4 Effect of MSQ Score

As MSQ scores were selected as one of the two features used for linear classification any variance in the function used to compute these score will significantly change the accuracy of the classifier. In section 6.2 the drawback on the effect of the implemented MSQ score on analysis was outlined and this effect may be a direct consequence of a sub-optimal selection of the parameters required by the two peak finding algorithms that together form the MSQ score.

As outlined in section 4.4.3 these parameters were selected by conducting a two dimensional grid search on the training data for the delta required by Billauer's peak finder and the distance required by the Scipy's peak finder that provided a low  $p$  value of the null hypothesis for both the Man-Whitney test for AB vs C segmentation and A vs BC segmentation. As the region of parameters that corresponded to the highest ability to classify the two different classification problems were different the final values were selected empirically corresponding to values that represented a good compromise between the segmentation abilities.

The results outlined and explained in 6.2 suggest that this selection was sub-optimal as it resulted in the undesirably close distribution of A class signals to C class signals. This is likely as a result of the uneven distribution of the class labels where as the majority of signals within the training set are of class B the use of the Man-Whitney U test for the metric of the two dimensional grid search was likely biased towards segmenting class B signals.

## 7.5 Optimal length Omission

As outlined in section 3.1.5 the development of the algorithms used for classification assumed a signal sample length of three seconds to be sufficient for classification, as suggested by Elgendi et Al in [56]. Although this was found to be sufficient no investigation was conducted to suggest such a sample length was in-fact optimal. This is because such an investigation would require all data to be manually segmented pulse wise such that an annotator would then have to annotate the quality of every pulse within a signal. This was considered infeasible within the time constraints of the project however such an investigation would provide valuable information on the trade off between the resolution of the algorithms developed and their accuracies.

## 7.6 Precision and Accuracy of Speed Tests

As outlined in section, when measuring the computational intensity of the different SQIs, the raw value of the time required is dependent on the processor used. Although these raw times are subject to change no decisions were made on raw times, instead the relative time of computation for different SQIs and algorithms were used in sections 4.4.4 and 6.3 for comparison and decision making instead. These relative computational times will not vary significantly across processors as the underlying complexity of the algorithms explained in section 6.4 and the required number of computations remain the same. The precision of these times was also ensured by conducting a large number of tests with a large number of repeats and was found to be sufficient considering the large changes in computational times observed.

As mentioned in section 6.4 the standard method for computing the auto-correlation of a signal sample was used when testing the speed of the auto-correlation based SQIs and they were

found found to have higher order computational complexity and hence require longer computation time. More efficient methods of finding the autocorrelation of a sample exist [105]–[107] and could increase the viability of auto-correlation based SQIs for inexpensive real-time classification.

## 7.7 Draw Backs of Universality

Significant effort was made to ensure the universality of the algorithm developed by selecting various datasets using different devices for PPG signal acquisition and with varying amounts of prepossessing. This was conducted such that to highlight the universal utility of the SQIs selected (MSQ and Skewness), as well as the linear classifier selected to provide thresholds for linear classification , however, this will limit the accuracy of any developed classifier for a given device. Although the accuracy remains sufficient for application, the intention for such universality was to illustrate and validate the proposed universal framework used for algorithm development such that for a given sensor and a sufficient amount of training data one could follow the framework highlighted by this report to extract thresholds corresponding to an increased accuracy for this device.

## 7.8 Effect of Annotation

An important investigation that would further validate the success of the framework and algorithm developed for universal classification is the manual annotation of the training data by other annotators. This is because although specific requirements were outlined to be followed during the annotation of signal samples, manual annotation remains to be subjective and was found to vary slightly between annotators in [56] and [57]. Considering the unanticipated shift to remote learning and the resulting time constraints for project development this was omitted from the investigation considering the amount of data that would have to be annotated by someone else.

## 7.9 Dataset Size and Distribution of the dataset

The effects of the distribution of the different classes within the dataset used for both training and testing are outlined in sections 4.3, 6.1.1 and 6.1.2. Although this distribution was found to be representative of the practical distribution of data the development of classifiers using machine learning methods will nevertheless benefit from more balanced datasets containing more samples of the minority classes. As outlined in the previous section the unanticipated shift to remote learning and the resulting time constraints for project development meant that annotation of more signals to attain a larger dataset was not possible.

## Chapter 8

# Conclusions and Further Work

To conclude the findings of this report 26 SQIs were evaluated on their ability to segment pre-processed three second samples, acquired from a sufficient range of PPG acquisition devices, into three different manually annotated classes. A real-time pre-processing technique was designed and implemented such that to allow for universal classification of the range of data acquired.

This SQI evaluation consisted of a Man-Whitney U and Wilcoxon signed-rank Test to asses segmentation independently, a Random Forest feature importance test to evaluate their importance in classification when used together as well as a computation intensity test where the relative computation times were assessed, an evaluation missing in classification development in the literature.

This evaluation identified skewness and MSQ SQIs as scores with sufficiently low computation time and sufficient segmentation ability, to be able to independently classify three second samples of a PPG signal both validating and expanding upon research conducted in [56], [57], [75]. An investigation into the optimal universal method of obtaining linear thresholds for classification was subsequently investigated. This investigation concluded that a linear classifier trained with minimum squared loss was found to robustly provide ideal thresholds for selection.

Auto-correlation based SQIs proposed in [57], [58], [75] were also found to have superior segmentation ability, however, were omitted from algorithm development as a result of their computational complexity. As mentioned in section 7.6 methods for more efficient computation of SQIs exist and could increase the applicability of auto-correlation based SQIs. This investigation is left as future work.

Both a three class and a two class algorithm were then developed using the extracted linear thresholds obtained by a square loss linear classifier. These algorithms were designed for real-time classification and were implemented into a graphic user interface to showcase its use in classifying the quality of three second samples acquired in real time by the AFE4900EVM as well as allow for the classification of pre recorded imported signals.

These algorithms were tested on their ability to segment a test set of data samples made up of all the subsets acquired to ensure universality and found to have a justifiable accuracy to claim universality of the framework used for algorithm development. The universality of the training data was found to limit the accuracy of said classification when classifying data specific to a given acquisition device. Tools were hence designed and created such that to utilise the framework for universal development of a classification algorithm to identify thresholds for optimal classification for a specific device.

An analysis and evaluation of the distribution of the training samples within the new two dimensional feature-space was conducted and a limitation of the implemented MSQ score's ability to classify class A signals was identified in section 7.4. A re-optimisation of the MSQ score with a training set with more equal distribution of the three classes and a weighted sum loss function comprising of AB vs C and A vs BC segmentation scores could significantly improve classification accuracy and was left as future work.

The need for validation of the universality of the algorithm through the analysis of its performance when subject to different annotation was also identified in section 7.8 and is recommended in any future work. A graphic user interface for simple visual annotation was developed for this purpose, the output of which can be imported into a jupyter notebook which allows for a revaluation of all the investigations conducted in chapter 4.

An investigation into the optimal signal sample length was also identified in section 7.5 to

provide valuable information on the trade off between the resolution of the algorithms developed and their accuracies, this was left as future work due to the time constraining pulse wise manual annotation required.

One could also improve the update rate of the classification by implementing a sliding window used as the input to the algorithm such that the input is always three seconds long however can be updated in much smaller time periods. A subsequent investigation on the trade off between update rate and computational intensity would also be valuable in future development. With the update rate reduced one could subsequently integrate this classification algorithm into a algorithm for autonomous calibration of ppg sensors and probes to a specific patient and probe placement, this would significantly improve the average quality of ppg signals acquired from all patients and significantly reduce the time required for probe development.

The algorithm was implemented in python as the AFE4900EVM driver being developed for ppg signal acquisition was written in python, however, in future possible wearable applications of the algorithm it is recommended to implement and optimise the algorithm in C or C++. This is because it would allow the possibility of the algorithm being performed in a separate thread of the micro-controller designed to acquire ppg signals from the AFE4900 independently, this would also allow for various valuable simple applications of the classification algorithm developed, for example, using LEDs or other methods to guide untrained users on the correct placement of the sensors.

# Chapter 9

## User Guide

A GitHub repository containing details of all the files and tools required to both validate the findings of this report and develop algorithms specific to an external or internal dataset can be found at [Here](#). The GitHub repository also contains the files required to run the GUI outlined in 5.1, details of the libraries required and general start-up guides for all tools and files can be found in the documentation of said repository.

# Bibliography

- [1] Scott B. Halstead. “Dengue: The Syndromic Basis to Pathogenesis Research. Inutility of the 2009 WHO Case Definition”. In: *The American Journal of Tropical Medicine and Hygiene* 88.2 (2013), pp. 212–215. ISSN: 0002-9637. DOI: <https://doi.org/10.4269/ajtmh.12-0197>. URL: <https://www.ajtmh.org/content/journals/10.4269/ajtmh.12-0197>.
- [2] Carolin Fleischmann, André Scherag, Neill Adhikari, et al. “Assessment of Global Incidence and Mortality of Hospital-treated Sepsis - Current Estimates and Limitations”. In: *American journal of respiratory and critical care medicine* 193 (Sept. 2015). DOI: 10.1164/rccm.201504-0781OC.
- [3] Mariska Weenk, Harry van Goor, Bas Frietman, et al. “Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study”. In: *JMIR mHealth and uHealth* 5.7 (July 2017), e91–e91.
- [4] Jeroen Ludikhuize, Susanne M Smorenburg, Sophia E de Rooij, et al. “Identification of deteriorating patients on general wards; measurement of vital parameters and potential effectiveness of the Modified Early Warning Score”. In: *Journal of critical care* 27.4 (2012), 424–e7.
- [5] Una Kyriacos, Jennifer Jelsma, and Sue Jordan. “Record review to explore the adequacy of post-operative vital signs monitoring using a local modified early warning score (mews) chart to evaluate outcomes”. In: *PLoS One* 9.1 (2014).
- [6] Matthew M Churpek, Trevor C Yuen, and Dana P Edelson. “Predicting clinical deterioration in the hospital: the impact of outcome selection”. In: *Resuscitation* 84.5 (2013), pp. 564–568.
- [7] Ila D Mapp, Leslie L Davis, and Heidi Krowchuk. “Prevention of unplanned intensive care unit admissions and hospital mortality by early warning systems”. In: *Dimensions of Critical Care Nursing* 32.6 (2013), pp. 300–309.
- [8] Sophie Yacoub and Bridget Wills. “Predicting outcome from dengue”. In: *BMC Medicine* 12.1 (2014), p. 147. DOI: 10.1186/s12916-014-0147-9. URL: <https://doi.org/10.1186/s12916-014-0147-9>.
- [9] Jean-Louis Vincent. “The Clinical Challenge of Sepsis Identification and Monitoring”. In: *PLoS medicine* 13.5 (May 2016), e1002022–e1002022.
- [10] Laure Wynants, Ben Van Calster, Gary S Collins, et al. “Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal”. In: *BMJ* 369 (2020). DOI: 10.1136/bmj.m1328. eprint: <https://www.bmjjournals.com/content/369/bmj.m1328.full.pdf>. URL: <https://www.bmjjournals.com/content/369/bmj.m1328>.
- [11] John Allen. “Photoplethysmography and its application in clinical physiological measurement”. In: *Physiol. Meas* 28 (Jan. 2007).
- [12] Greeshma Joseph, Almaria Joseph, Geevarghese Titus, et al. “Photoplethysmogram (PPG) signal analysis and wavelet de-noising”. In: *2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMRD)* (2014), pp. 1–5.
- [13] Lewis Mackenzie. “In vivo microvascular oximetry using multispectral imaging”. PhD thesis. Durham University, Nov. 2016. DOI: 10.13140/RG.2.2.31384.80649.
- [14] Deric P Jones. “Medical electro-optics: measurements in the human microcirculation”. In: *Physics in Technology* 18.2 (Mar. 1987), pp. 79–85. DOI: 10.1088/0305-4624/18/2/305. URL: <https://doi.org/10.1088%2F0305-4624%2F18%2F2%2F305>.

- [15] GORDY E and DRABKIN DL. "Spectrophotometric studies. XVI Determination of the oxygen saturation of blood by a simplified technique, applicable to standard equipment". In: *Journal of biological chemistry*. 227 (1957), p. 285. ISSN: 0021-9258.
- [16] A Murray and D Marjanovic. "Optical assessment of recovery of tissue blood supply after removal of externally applied pressure". In: *Medical & biological engineering & computing*. 35.4 (1997), pp. 425–427. ISSN: 0140-0118.
- [17] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, et al. "A review on wearable photoplethysmography sensors and their potential future applications in health care". In: *International journal of biosensors & bioelectronics* 4.4 (2018), pp. 195–202.
- [18] Seulki Ha, Sangheon Park, Hyeoncheol Lim, et al. "The placement position optimization of a biosensor array for wearable healthcare systems". In: *Journal of Mechanical Science and Technology* 33.7 (2019), pp. 3237–3244.
- [19] P. Renevey, J. Solà, P. Theurillat, et al. "Validation of a wrist monitor for accurate estimation of RR intervals during sleep". In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013, pp. 5493–5496.
- [20] Lena Nilsson, Anders Johansson, and Sigga Kalman. "Respiration can be monitored by photoplethysmography with high sensitivity and specificity regardless of anaesthesia and ventilatory mode". In: *Acta anaesthesiologica scandinavica* 49.8 (2005), pp. 1157–1162.
- [21] Björn Jönsson, Claes Laurent, Tommy Skau, et al. "A new probe for ankle systolic pressure measurement using photoplethysmography (PPG)". In: *Annals of biomedical engineering* 33.2 (2005), pp. 232–239.
- [22] Kirk H Shelley, Doris Tamai, Denis Jablonka, et al. "The effect of venous pulsation on the forehead pulse oximeter wave form as a possible source of error in Spo<sub>2</sub> calculation". In: *Anesthesia & Analgesia* 100.3 (2005), pp. 743–747.
- [23] D. B. McCombie, A. T. Reisner, and H. H. Asada. "Adaptive blood pressure estimation from wearable PPG sensors using peripheral artery pulse wave velocity measurements and multi-channel blind identification of local arterial dynamics". In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. Aug. 2006, pp. 3521–3524. DOI: 10.1109/IEMBS.2006.260590.
- [24] Joseph M. Schmitt, Guan-Xiong Zhou, and Justin Miller. "Measurement of blood hematocrit by dual-wavelength near-IR photoplethysmography". In: *Physiological Monitoring and Early Detection Diagnostic Methods*. Ed. by Thomas S. Mang. Vol. 1641. International Society for Optics and Photonics. SPIE, 1992, pp. 150–161. DOI: 10.1117/12.59360. URL: <https://doi.org/10.1117/12.59360>.
- [25] Shota Ekuni and Yoshiyuki Sankai. "Non-invasive measurement method of blood hematocrit for management of extracorporeal circulation". In: *2014 IEEE/SICE International Symposium on System Integration, SII 2014* (Jan. 2015), pp. 76–81. DOI: 10.1109/SII.2014.7028015.
- [26] I. García-López, P. Sharma, and E. Rodriguez-Villegas. "Heart Rate Extraction from Novel Neck Photoplethysmography Signals". In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019, pp. 6541–6544.
- [27] Scott B. Halstead. "Dengue: The Syndromic Basis to Pathogenesis Research. Inutility of the 2009 WHO Case Definition". In: *The American Journal of Tropical Medicine and Hygiene* 88.2 (2013), pp. 212–215. ISSN: 0002-9637. DOI: <https://doi.org/10.4269/ajtmh.12-0197>. URL: <https://www.ajtmh.org/content/journals/10.4269/ajtmh.12-0197>.
- [28] T P Monath. "Dengue: the risk to developed and developing countries". In: *Proceedings of the National Academy of Sciences* 91.7 (1994), pp. 2395–2400. ISSN: 0027-8424. DOI: 10.1073/pnas.91.7.2395. eprint: <https://www.pnas.org/content/91/7/2395.full.pdf>. URL: <https://www.pnas.org/content/91/7/2395>.
- [29] Daniel S Kassavini, Yen-Hong Kuo, and Nasim Ahmed. "Initial systolic blood pressure and ongoing internal bleeding following torso trauma". In: *Journal of emergencies, trauma, and shock* 4.1 (Jan. 2011), pp. 37–41.

- [30] G. Wang, M. Atef, and Y. Lian. "Towards a Continuous Non-Invasive Cuffless Blood Pressure Monitoring System Using PPG: Systems and Circuits Review". In: *IEEE Circuits and Systems Magazine* 18.3 (July 2018), pp. 6–26. ISSN: 1558-0830. DOI: 10.1109/MCAS.2018.2849261.
- [31] Q. Zhang, X. Zeng, W. Hu, et al. "A Machine Learning-Empowered System for Long-Term Motion-Tolerant Wearable Monitoring of Blood Pressure and Heart Rate With Ear-ECG/PPG". In: *IEEE Access* 5 (2017), pp. 10547–10561.
- [32] J. Crighton Bramwell and Archibald Vivian Hill. "The velocity of pulse wave in man". In: *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* 93.652 (1922), pp. 298–306. DOI: 10.1098/rspb.1922.0022. eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.1922.0022>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1922.0022>.
- [33] D.J. Hughes, C.F. Babbs, L.A. Geddes, et al. "Measurements of Young's modulus of elasticity of the canine aorta with ultrasound". In: *Ultrasonic Imaging* 1.4 (1979), pp. 356–367. ISSN: 0161-7346. DOI: [https://doi.org/10.1016/0161-7346\(79\)90028-2](https://doi.org/10.1016/0161-7346(79)90028-2). URL: <http://www.sciencedirect.com/science/article/pii/0161734679900282>.
- [34] Yont Joo KWON, Min KANG, Sun Kwon KIM, et al. *BLOOD PRESSURE MEASURING USING APPARATUS, AND BLOOD PRESSURE MEASURING APPARATUS USING LIGHT SOURCE SELECTION PROCESS*.
- [35] Texas Instruments. *Data-sheet: AFE4900 Ultra-lowPower, Integrated AFE for Wearable Optical, Electrical Bio-sensing with FIF*. Tech. rep. Texas Instruments, 2019.
- [36] OSRAM Opto Semiconductors. *Data-sheet: SFH7072*. Tech. rep. OSRAM Opto Semiconductors, 2018.
- [37] OSRAM Opto Semiconductors. *Data-sheet: SFH7050*. Tech. rep. OSRAM Opto Semiconductors, 2016.
- [38] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, et al. "A review on wearable photoplethysmography sensors and their potential future applications in health care". In: *International journal of biosensors & bioelectronics* 4.4 (2018), pp. 195–202.
- [39] Mathieu Lemay, Mattia Bertschi, Josep Solà, et al. "Application of Optical Heart Rate Monitoring". In: Dec. 2014, pp. 105–129. DOI: 10.1016/B978-0-12-418662-0.00023-4.
- [40] YongKwi Lee, HyunSoon Shin, Jun Jo, et al. "Development of a wristwatch-type PPG array sensor module". In: *2011 IEEE International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*. IEEE. 2011, pp. 168–171.
- [41] Edward Sazonov. *Wearable Sensors: Fundamentals, implementation and applications*. Elsevier, 2014.
- [42] C. Orphanidou, T. Bonnici, P. Charlton, et al. "Signal-Quality Indices for the Electrocardiogram and Photoplethysmogram: Derivation and Applications to Wireless Monitoring". In: *IEEE Journal of Biomedical and Health Informatics* 19.3 (2015), pp. 832–838.
- [43] Vishal Nangalia, David R Prytherch, and Gary B Smith. "Health technology assessment review: Remote monitoring of vital signs-current status and future challenges". In: *Critical Care* 14.5 (2010), p. 233.
- [44] The Joint Commision. *Sentinel Event Alert 50: Medical device alarm safety in hospitals*. Tech. rep. The Joint Commision, 2013.
- [45] ECRI. *Alarm Safety Handbook*. ECRI, 2014.
- [46] J Abdul Sukor, SJ Redmond, and NH Lovell. "Signal quality measures for pulse oximetry through waveform morphology analysis". In: *Physiological measurement* 32.3 (2011), p. 369.
- [47] Emad Kasaayan Naeini, Iman Azimi, Amir M. Rahmani, et al. "A Real-time PPG Quality Assessment Approach for Healthcare Internet-of-Things". In: *Procedia Computer Science* 151 (2019). The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops, pp. 551–558. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2019.04.074>. URL: <http://www.sciencedirect.com/science/article/pii/S1877050919305368>.

- [48] Maik Pflugradt, Benjamin Moeller, and Reinhold Orglmeister. “OPRA: A fast on-line signal quality estimator for pulsatile signals”. In: *IFAC-PapersOnLine* 48.20 (2015). 9th IFAC Symposium on Biological and Medical Systems BMS 2015, pp. 459–464. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2015.10.183>. URL: <http://www.sciencedirect.com/science/article/pii/S2405896315020741>.
- [49] Wilhelm von Rosenberg, Theerasak Chanwimalueang, Valentin Goverdovsky, et al. “Hearables: Feasibility of recording cardiac rhythms from head and in-ear locations”. In: *Royal Society Open Science* 4 (Nov. 2017), p. 171214. DOI: 10.1098/rsos.171214.
- [50] X. Sun, P. Yang, and Y. Zhang. “Assessment of photoplethysmogram signal quality using morphology integrated with temporal information approach”. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012, pp. 3456–3459.
- [51] Q Li and G D Clifford. “Dynamic time warping and machine learning for signal quality assessment of pulsatile signals”. In: *Physiological Measurement* 33.9 (Aug. 2012), pp. 1491–1501. DOI: 10.1088/0967-3334/33/9/1491. URL: <https://doi.org/10.1088%2F0967-3334%2F33%2F9%2F1491>.
- [52] Elyas Sabeti, Narathip Reamaroon, Michael Mathis, et al. “Signal quality measure for pulsatile physiological signals using morphological features: Applications in reliability measure for pulse oximetry”. In: *Informatics in Medicine Unlocked* 16 (2019), p. 100222. ISSN: 2352-9148. DOI: <https://doi.org/10.1016/j.imu.2019.100222>. URL: <http://www.sciencedirect.com/science/article/pii/S2352914819301856>.
- [53] W Karlen, K Kobayashi, J M Ansermino, et al. “Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation”. In: *Physiological Measurement* 33.10 (Sept. 2012), pp. 1617–1629. DOI: 10.1088/0967-3334/33/10/1617. URL: <https://doi.org/10.1088%2F0967-3334%2F33%2F10%2F1617>.
- [54] G. B. Papini, P. Fonseca, X. L. Aubert, et al. “Photoplethysmography beat detection and pulse morphology quality assessment for signal reliability estimation”. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2017, pp. 117–120.
- [55] Tania Pereira, Kais Gadhouni, Mitchell Ma, et al. “A supervised approach to robust photoplethysmography quality assessment”. In: *IEEE Journal of Biomedical and Health Informatics* (2019).
- [56] Mohamed Elgendi. “Optimal Signal Quality Index for Photoplethysmogram Signals”. In: *Bioengineering* 3.4 (Sept. 2016), p. 21. ISSN: 2306-5354. URL: <http://dx.doi.org/10.3390/bioengineering3040021>.
- [57] Nikhilesh Pradhan. “Evaluation of the Signal Quality of Wrist-Based Photoplethysmography”. PhD thesis. Carleton University, 2018.
- [58] S. Vadrevu and M. S. Manikandan. “Real-Time PPG Signal Quality Assessment System for Improving Battery Life and False Alarms”. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 66.11 (2019), pp. 1910–1914.
- [59] N. Selvaraj, Y. Mendelson, K. H. Shelley, et al. “Statistical approach for the detection of motion/noise artifacts in Photoplethysmogram”. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011, pp. 4972–4975.
- [60] Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing*. 3rd. USA: Prentice Hall Press, 2009.
- [61] L. Bluestein. “A linear filtering approach to the computation of discrete Fourier transform”. In: *IEEE Transactions on Audio and Electroacoustics* 18.4 (1970), pp. 451–455.
- [62] *Butterworth Filter Design and Low Pass Butterworth Filters*. July 2018. URL: [https://www.electronics-tutorials.ws/filter/filter\\_8.html](https://www.electronics-tutorials.ws/filter/filter_8.html).
- [63] I. W. Selesnick and C. S. Burrus. “Generalized digital Butterworth filter design”. In: *IEEE Transactions on Signal Processing* 46.6 (1998), pp. 1688–1694.
- [64] R. Krishnan, B. Natarajan, and S. Warren. “Two-Stage Approach for Detection and Reduction of Motion Artifacts in Photoplethysmographic Data”. In: *IEEE Transactions on Biomedical Engineering* 57.8 (2010), pp. 1867–1876.

- [65] Nandakumar Selvaraj, Yitzhak Mendelson, Kirk Shelley, et al. “Statistical approach for the detection of motion/noise artifacts in Photoplethysmogram”. In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2011* (Aug. 2011), pp. 4972–5. DOI: 10.1109/IEMBS.2011.6091232.
- [66] Shanbao Tong, Zhengjun Li, Yisheng Zhu, et al. “Describing the Nonstationarity Level of Neurological Signals Based on Quantifications of Time–Frequency Representation”. In: *Biomedical Engineering, IEEE Transactions on* 54 (Nov. 2007), pp. 1780–1785. DOI: 10.1109/TBME.2007.893497.
- [67] R. R. Coifman and M. V. Wickerhauser. “Entropy-based algorithms for best basis selection”. In: *IEEE Transactions on Information Theory* 38.2 (1992), pp. 713–718.
- [68] C. H. Chen. “Signal processing handbook”. In: *Signal processing handbook*. 1988.
- [69] Saeed V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.
- [70] Mohamed Elgendi, Ian Norton, Matt Brearley, et al. “Systolic Peak Detection in Acceleration Photoplethysmograms Measured from Emergency Responders in Tropical Conditions”. In: *PloS one* 8 (Oct. 2013), e76585. DOI: 10.1371/journal.pone.0076585.
- [71] Bing Nan Li, Ming Chui Dong, and Mang I. Vai. “On an automatic delineator for arterial blood pressure waveforms”. In: *Biomedical Signal Processing and Control* 5.1 (2010), pp. 76–81. ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2009.06.002>. URL: <http://www.sciencedirect.com/science/article/pii/S1746809409000470>.
- [72] Eli Billauer. *peakdet: Peak detection using MATLAB*. 2012. URL: <http://billauer.co.il/peakdet.html>.
- [73] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: <https://doi.org/10.1038/s41592-019-0686-2>.
- [74] P. Welch. “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms”. In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), pp. 70–73.
- [75] Nikhilesh Pradhan, Sreeraman Rajan, and Andy Adler. “Evaluation of the signal quality of wrist-based photoplethysmography”. In: *Physiological Measurement* 40.6 (July 2019), p. 065008. DOI: 10.1088/1361-6579/ab225a. URL: <https://doi.org/10.1088%2F1361-6579%2Fab225a>.
- [76] Jeremiah Wander and D Morris. “A combined segmenting and non-segmenting approach to signal quality estimation for ambulatory photoplethysmography”. In: *Physiological measurement* 35 (Dec. 2014), pp. 2543–61. DOI: 10.1088/0967-3334/35/12/2543.
- [77] N. Pradhan, S. Rajan, A. Adler, et al. “Classification of the quality of wristband-based photoplethysmography signals”. In: *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2017, pp. 269–274.
- [78] H. B. Mann and D. R. Whitney. “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other”. In: *Ann. Math. Statist.* 18.1 (Mar. 1947), pp. 50–60. DOI: 10.1214/aoms/1177730491. URL: <https://doi.org/10.1214/aoms/1177730491>.
- [79] Michael P Fay and Michael A Proschan. “Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules”. In: *Statistics surveys* 4 (2010), pp. 1–39.
- [80] Frank Wilcoxon. “Individual Comparisons by Ranking Methods”. In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. ISSN: 00994987. URL: <http://www.jstor.org/stable/3001968>.
- [81] Stacey Ronaghan. *The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark*. Nov. 2019. URL: <https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3>.
- [82] David Powers and Ailab. “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation”. In: *J. Mach. Learn. Technol* 2 (Jan. 2011), pp. 2229–3981. DOI: 10.9735/2229-3981.

- [83] Koo P Shung. *Accuracy, Precision, Recall or F1?* 2020. URL: <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>.
- [84] Daniel Berrar. “Cross-Validation”. In: Jan. 2018. DOI: 10.1016/B978-0-12-809633-8.20349-X.
- [85] Robert Berwick. *An Idiot’s guide to Support vector machines (SVMs)*. 2011. URL: <http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf>.
- [86] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. USA: Cambridge University Press, 2008.
- [87] Robert Berwick. *SVMs & Boosting, II*. 2011. URL: <http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf>.
- [88] Czako Zoltan. *SVM and Kernel SVM*. May 2020. URL: <https://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200>.
- [89] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [90] P. Charbonnier, L. Blanc-Feraud, G. Aubert, et al. “Deterministic edge-preserving regularization in computed imaging”. In: *IEEE Transactions on Image Processing* 6.2 (1997), pp. 298–311.
- [91] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [92] C. E. Shannon. “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.4 (1948), pp. 623–656. DOI: 10.1002/j.1538-7305.1948.tb00917.x. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1948.tb00917.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1948.tb00917.x>.
- [93] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, et al. “Toward a Robust Estimation of Respiratory Rate From Pulse Oximeters”. In: *IEEE Transactions on Biomedical Engineering* 64.8 (2017), pp. 1914–1923.
- [94] Mohammed Saeed, Mauricio Villarroel, Andrew T Reisner, et al. “Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database”. In: *Critical care medicine* 39.5 (May 2011), pp. 952–960. DOI: 10.1097/CCM.0b013e31820a92c6. URL: <https://pubmed.ncbi.nlm.nih.gov/21283005>.
- [95] Ary L Goldberger, Luis AN Amaral, Leon Glass, et al. “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals”. In: *Circulation* 101.23 (2000), e215–e220.
- [96] *Smartcare Devices*. <http://devices.smartcareanalytics.co.uk/>. 2016.
- [97] Hartmut Gehring, Christoph Hornberger, Holger Matz, et al. “The effects of motion artifact and low perfusion on the performance of a new generation of pulse oximeters in volunteers undergoing hypoxemia”. In: *Respiratory care* 47 (Jan. 2002), pp. 48–60.
- [98] Maxime Cannesson, Bertrand Delannoy, Antoine Morand, et al. “Does the Pleth Variability Index Indicate the Respiratory-Induced Variation in the Plethysmogram and Arterial Pressure Waveforms?” In: *Anesthesia and analgesia* 106 (Apr. 2008), 1189–94, table of contents. DOI: 10.1213/ane.0b013e318167ab1f.
- [99] Douglas A. Colquhoun, Katherine T. Forkin, Marcel E. Durieux, et al. “Ability of the Masimo pulse CO-Oximeter to detect changes in hemoglobin”. In: *Journal of Clinical Monitoring and Computing* 26.2 (2012), pp. 69–73.
- [100] Caifeng Shan. *Patent for Motion robust vital signal monitoring*.
- [101] Guido Van Rossum. *The Python Library Reference, release 3.8.2*. Python Software Foundation, 2020.
- [102] David Beazley and Brian K. Jones. *Python Cookbook*. O’Reilly Media, Inc., 2013.
- [103] Lorenzo Rosasco, Ernesto De Vito, Andrea Caponnetto, et al. “Are Loss Functions All the Same?” In: *Neural Comput.* 16.5 (May 2004), pp. 1063–1076. ISSN: 0899-7667. DOI: 10.1162/089976604773135104. URL: <https://doi.org/10.1162/089976604773135104>.

- [104] PyQt. “PyQt Reference Guide”. In: (2012). URL: <http://www.riverbankcomputing.com/static/Docs/PyQt4/html/index.html>.
- [105] George Box, Gwilym Jenkins, and C. Reinsel. “Time series analysis: forecasting and control (third ed)”. In: *Time Series Analysis: Forecasting and Control: Fourth Edition* (May 2013). DOI: 10.1002/9781118619193.
- [106] “Appendix F - Saving CPU Time”. In: *Understanding Molecular Simulation (Second Edition)*. Ed. by Daan Frenkel and Berend Smit. Second Edition. San Diego: Academic Press, 2002, pp. 545–558. ISBN: 978-0-12-267351-1. DOI: <https://doi.org/10.1016/B978-012267351-1/50025-0>. URL: <http://www.sciencedirect.com/science/article/pii/B9780122673511500250>.
- [107] Peter H. Colberg and Felix Höfling. “Highly accelerated simulations of glassy dynamics using GPUs: Caveats on limited floating-point precision”. In: *Computer Physics Communications* 182.5 (May 2011), pp. 1120–1129. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2011.01.009. URL: <http://dx.doi.org/10.1016/j.cpc.2011.01.009>.

# Appendix A

## Figures

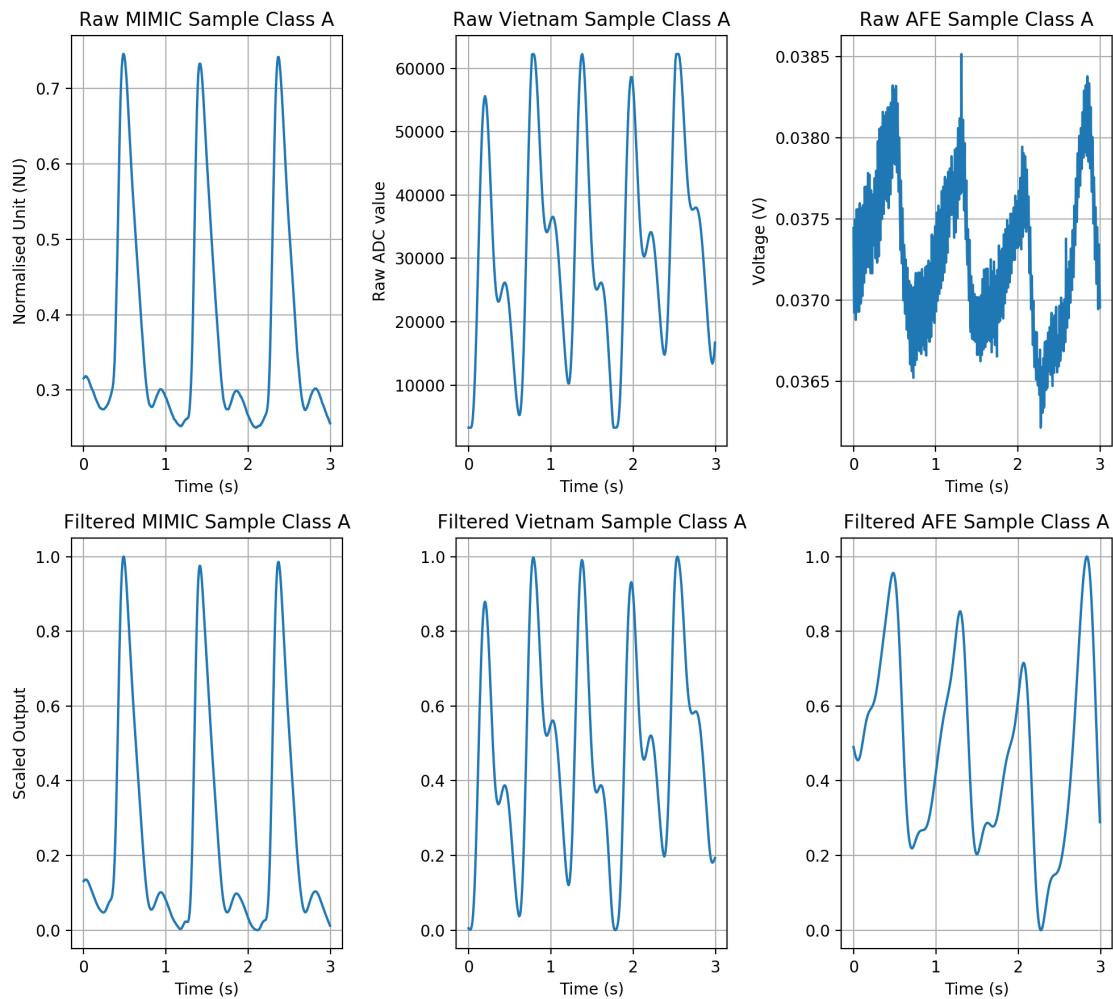


Figure A.1: Examples of class A samples from each subset raw and filtered

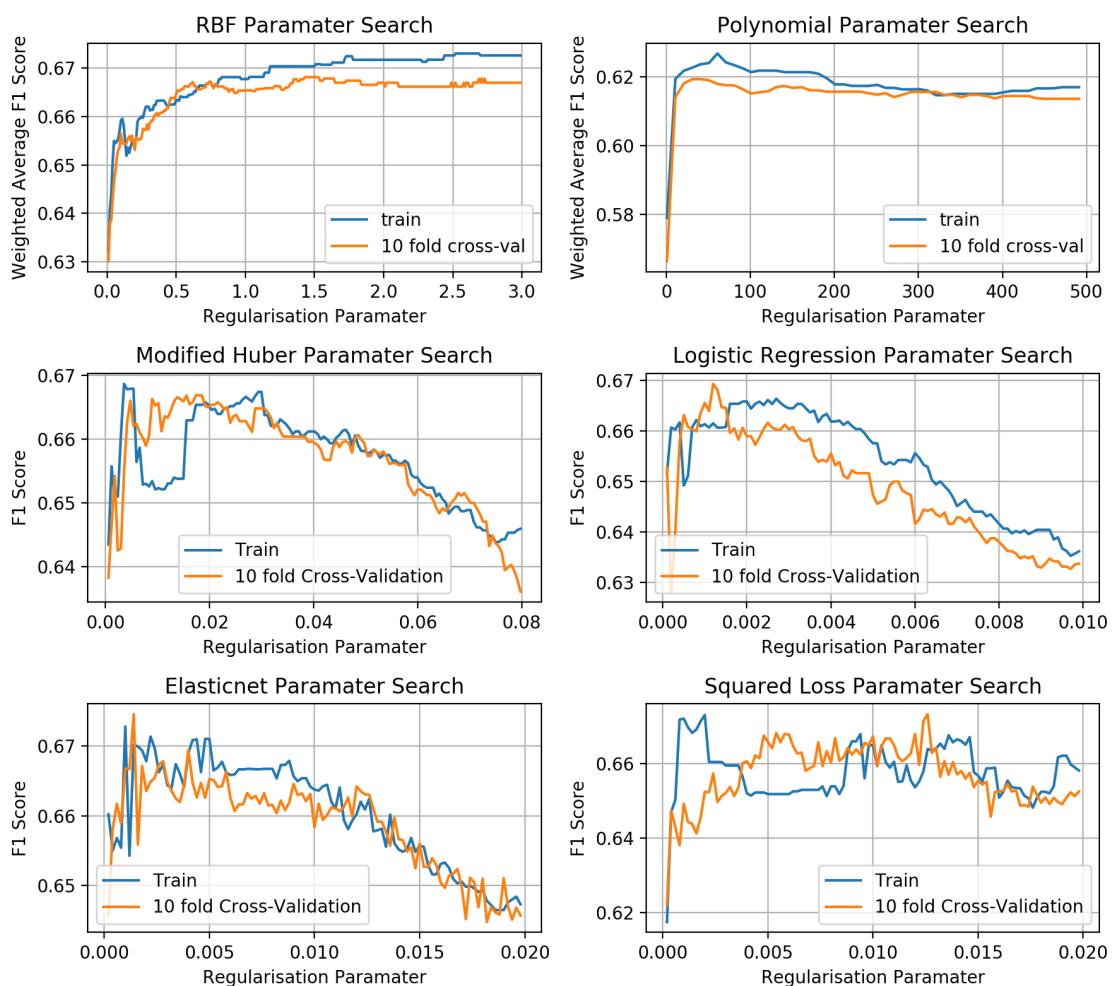


Figure A.2: Regularisation Parameter optimisation for varying three class classifiers

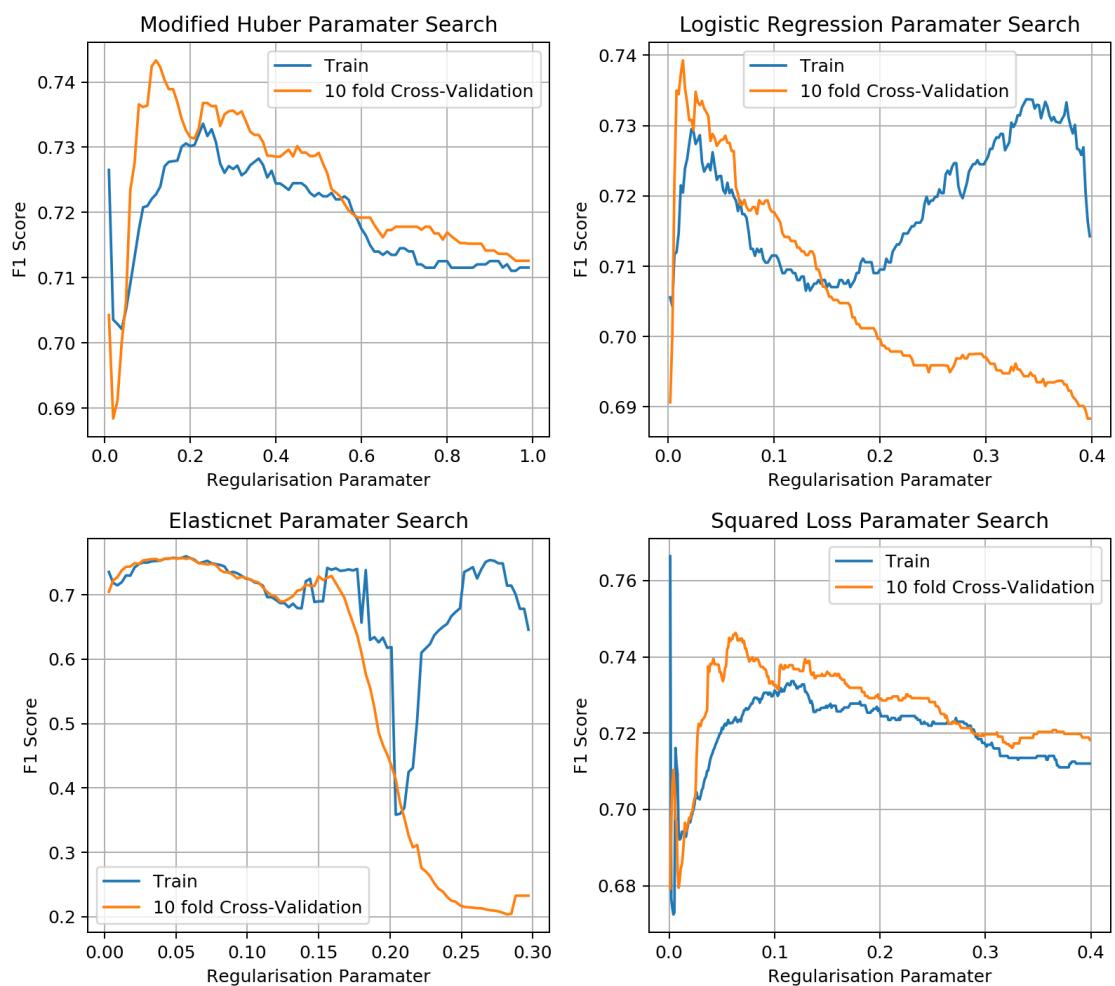


Figure A.3: Regularisation Parameter optimisation for varying two