Development of a mobile application for metabolic syndrome screening

JM Crause

orcid.org/0000-0001-8708-5323

Dissertation accepted in fulfilment of the requirements for the degree *Master of Engineering in Computer and Electronic Engineering* at the North-West University

Supervisor: Prof PA van Vuuren

Graduation: May 2022

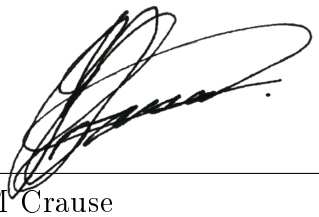
Student number: 24117986

Declaration

I, Jacobus Mayer Crause, declare that this dissertation, entitled "Development of a mobile application for metabolic syndrome screening", is a presentation of my own original work. Whenever contributions of others are involved, every effort was made to indicate this clearly, with due reference to the literature.

No part of this work has been submitted in the past, or is being submitted, for a degree or examination at any other university or course.

Signed on this, February 14, 2022, in Potchefstroom.



JM Crause

Acknowledgements

Prof Pieter, thank you for the guidance you have given me over the course of my studies. And particularly thank you for the patience you showed when my personal and work schedules did not always allow for much progress.

Special thank you to **Wernich** and **Almari**, who assisted me during the data collection process. At a moment's notice, with a tight schedule, asking for nothing in return, you helped me when I needed it. And I am incredibly grateful.

Further, thank you to **Prof Carina** and **Sr Adele** from the HART institute for providing a sphygmomanometer for testing, and to **Michelle** from Cachet Park for providing the location for the data collection.

I want to acknowledge my mother, **Corrie**, who contributed by printing and laminating "Thank you" cards for research participants and providing her home and support when I needed to finalise this dissertation. Also, thank you to my brother, **Wimpie**, for helping to proofread the dissertation.

Finally, thank you to all 133 participants who allowed me to measure their blood pressure and use their data in my research. Without your input this would not have been possible.

Abstract

This dissertation was created in pursuit of determining the viability of a smartphone based technique for metabolic syndrome screening. This screening method consists of two artificial neural networks. The first network uses a photoplethysmography (PPG) waveform measured using a smartphone camera and its FFT analysis to determine blood pressure. The second classifier uses the heart rate and blood pressure determined from the PPG measurement and user data, including sex, age, BMI, waist-to-height ratio, smoking, drinking habits, medical history and activity level, to determine whether a user may be at risk of metabolic syndrome.

The blood pressure network is initially trained and tested using PPG data, measured with a blood oximeter, from the MIMIC III online waveform dataset. The algorithm is then tested with PPG data obtained from participants with a smartphone camera. In real world scenarios the BP (blood pressure) algorithm has a mean absolute error of 7.7 mmHg ($\pm$6.9) for SBP (systolic blood pressure) and 8.0 mmHg ($\pm$6.3) for DBP (diastolic blood pressure).

Multiple metabolic syndrome algorithms are tested, with and without lifestyle related factors, and with a simulated error being added to blood pressure inputs based on results from participant tests. It was found that the inclusion of lifestyle factors improved the performance of the model by approximately 2%, from 75.2% to 77.2%, when assessed using 10-fold cross validation. When including simulated HR and BP errors, the network including lifestyle factors achieved an average accuracy of 74.9%. An approach similar to bootstrapping was followed where the 10-cross validation process is executed 10 times, randomly shuffling the data with each cycle, resulting in a total of 100 testing sets. The average accuracy across the 100 sets for the first model was 73.4%, with a sensitivity of 73.1% and a specificity of 73.6%. For the model with lifestyle factors the accuracy was 76.3%, with a sensitivity of 74.9% and a specificity of 77.4%. A t-test was done to show that the accuracy difference between the models was statistically significant ($t = -2.48$; $p < 0.014$).

A bias in the BP algorithm was found for the real world testing, causing the mean SBP to be 5 mmHg too low. Due to a strong dependence on DBP, this caused the MetS algorithm to perform less optimally. This highlights the necessity of an accurate BP algorithm for the MetS algorithm to provide reliable results. The accuracy obtained across the 100 testing sets when including HR and BP errors was 73.1%, with a sensitivity of 64.9% and a specificity of 79.9%.

In order to determine if the output from the model could be interpreted as a risk indicator instead of just a binary prediction, the models were assessed at different confidence levels. The models achieved 86.8% to 88.4% accuracy for confidence levels over 80% and 92.7% to 94.1% accuracy for confidence over 90%. This seems to indicate that there is definite value in using this model as a risk indicator, which a user could potentially use to determine whether a full diagnosis is warranted. In that sense, it falls perfectly in line with what is expected from a screening solution.

The results show that smartphone based metabolic syndrome screening is a promising venture, that can be expanded upon in further research and considered for real world screening solutions. However, further research would be required to test whether this solution achieves similar results across multiple devices before it could be rolled out as a final product.

Keywords:

Blood pressure, FFT, Metabolic Syndrome, Neural network, Photoplethysmography, Smartphone

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AAMI Association for the Advancement of Medical Instrumentation

ABP Arterial Blood Pressure

ANN Artificial Neural Network

ANSI American National Standards Institute

BMI Body Mass Index

BMR Basal Metabolic Rate

BP Blood Pressure

CDC Centers for Disease Control and Prevention

DBP Diastolic Blood Pressure

EGIR European Group for the study of Insulin Resistance

ESH European Society of Hypertension

FBDG Food-based dietary guidelines for South Africa

FFT Fast Fourier Transform

FNN Fuzzy Neural Network

HDL High-Density Lipoprotein

HIV Human Immunodeficiency Virus

HR Heart Rate

IDF International Diabetes Federation

ISO International Organization for Standardization

MAE Mean Absolute Error

MAPE Mean Absolute Percent Error

ME Mean Error

MetS Metabolic Syndrome

MIMIC Multiparameter Intelligent Monitoring in Intensive Care

MIT Massachusetts Institute of Technology

mNPV Modified Normalized Pulse Volume

MSE Mean Squared Error

NCEP ATPIII National Cholesterol Education Program Adult Treatment Panel III

PCA Principal Component Analysis

POC Proof of Concept

PPG Photoplethysmography

SBP Systolic Blood Pressure

SVM Support Vector Machine

SUS System Usability Scale

TB Tuberculosis

TEE Total Energy Expenditure

TPR Total Periphery Resistance

USBR United States Department of the Interior Bureau of Reclamation

WBS Work Breakdown Structure

WHO World Health Organisation

CHAPTER

1

INTRODUCTION

In this chapter we will explore the motivation for wanting to study techniques for screening metabolic syndrome. Specifically with the use of a mobile application. We'll discuss some the background of the problem through other relevant research in order to come to a research question.

1.1 Background

MetS[1] is a combination of risk factors that increases a person's risk of cardiovascular disease, type-2 diabetes mellitus, other associated metabolic diseases and mortality [1]–[3]. Subjects with MetS have a three-fold risk of a heart attack or stroke, two-fold risk of dying from cardiovascular disease, and five-fold greater risk of developing type-2 diabetes in both genders when compared to people without it [4]. There has also been some research that indicated MetS is associated with the development of severe COVID-19 [5].

Different organisations have defined MetS in slightly different manners. Irrespective of the definition used, MetS could usually be diagnosed with at least three of the following medical risk factors [1], [3], [4], [6]:

- abdominal obesity
- high fasting blood sugar

[1] Metabolic Syndrome

- high serum triglycerides
- low HDL[2] cholesterol
- high blood pressure.

Some definitions require one specific component in order to make a diagnosis, such as the IDF[3] criteria for which abdominal obesity is a requirement [6]. According to the NCEP ATPIII[4] criteria the combination of any three components could give a positive diagnosis [6]. The specific way in which a component is measured or determined and the values specified by each criteria may also differ slightly [6].

Since MetS is often associated with the harmful effects of weight gain, sedentary lifestyle, and/or an atherogenic diet, treating MetS will in most cases simply entail a change in these lifestyle habits. For those with a high fasting glucose level, there is evidence that a high fiber, low saturated fat diet with increased daily exercise can reduce the incidence of diabetes by nearly 60% [7]. Though intensified therapeutic lifestyle change will help the abnormal triglyceride and cholesterol levels, some patients may still require drug therapy, which is why it is recommended that those who suspect that they may have these conditions consult a medical doctor [7].

Globally MetS has reached almost epidemic proportions. In South Africa the picture is equally grim. Erasmus [8] showed MetS rates of more than 60% in a Bellville coloured population, also indicating that at this rate cardiovascular disease may reach epidemic status in the near future in South Africa. This is also made clear when considering the top ten causes of death in South Africa in 2018 as listed by the CDC[5] [9]:

1. TB
2. Diabetes
3. Cerebrovascular diseases
4. Other forms of heart disease
5. HIV
6. Influenza & Pneumonia
7. Hypertensive diseases
8. Other viral diseases
9. Chronic lower respiratory diseases

[2]High-Density Lipoprotein
[3]International Diabetes Federation
[4]National Cholesterol Education Program Adult Treatment Panel III
[5]Centers for Disease Control and Prevention

10. Ischaemic Health Disease[6].

From the list above five conditions (diabetes, cerebrovascular diseases, heart disease, hypertensive diseases and ischaemic health diseases) are related to metabolic syndrome. Erasmus [8] also found that nearly 20% of type 2 diabetes cases were undiagnosed. This shows that the high prevalence of MetS may be due to a lack of awareness. This further highlights the need for screening and treatment of MetS in South Africa.

Considering the lack of awareness of MetS and that the condition can be largely prevented and reversed with dietary, exercise and other lifestyle changes, a mobile application may be a good way to increase awareness of MetS. A screening tool would also be valuable as an indicator for further diagnosis. It has also been shown by [10] and [11] that employing computer aided diagnostic systems as a "second opinion" has led to improved diagnostic decisions, in diagnosis of related diseases like diabetes and pre-diabetes. This shows that an application providing sufficient feedback may help to lead users to better choices regarding the treatment or prevention of MetS.

In the future a tool like this could also serve as a guideline, giving users practical advice based on medical literature as well as national and international dietary and health guidelines, such as the WHO[7] guidelines or FBDG[8]. However, this is outside the scope of this dissertation.

One significant classification tool that will be used to perform this screening is machine learning, more specifically ANNs[9], using several non-invasive variables collected from the user, to identify the possibility of MetS.

1.2 Literature review

1.2.1 Metabolic Syndrome screening methods

Regardless of the definition being used, the presence of MetS significantly increases the long term risk for type 2 diabetes, cardiovascular disease, heart attack and stroke [1]–[3]. Kaur [4] conducted a study to assess age, gender, social status, employment, education, family history, physical activity, dietary habits, alcohol, sleep, body mass index and stress as determinants of MetS. This study showed correlation in most of these areas, and stated that *"common concerns of female gender, increasing age and BMI, sedentary lifestyle, stress and positive family history should be considered for early identification and appropriate intervention to fight the growing MetS epidemic"* [4]. Since all the items in the aforementioned study can easily be filled in on a questionnaire, this may be an easy way to gather information needed to screen for MetS.

[6] Also known as coronary artery disease
[7] World Health Organisation
[8] Food-based dietary guidelines for South Africa
[9] Artificial Neural Networks

There have been several studies using machine learning to identify subjects with MetS. An example of this is a study by Ushida [12], where an FNN[10] was used to search for combinational MetS risk factors. In this sudy the FNN found that a combination of an elevated level of γ-glutamyltranspeptidase (γ-GTP) and an elevated white blood cell count as the most significant combination of risk factors for the development of metabolic syndrome, where these factors highly correlated with alcohol and cigarette use respectively [12]. In a similar study by Park [13], a Bayesian network was used for predicting MetS using evolutionary attribute ordering. Another study by Worachartcheewan [14] used a decision tree for identification of MetS from the standard features, which reached an accuracy of 99.8% in diagnosing MetS. The research in [14] may also be of a lot of value for determining the relationship between the different components of MetS. A promising study was conducted by Ivanović [15], where an ANN was used to diagnose MetS using only non-invasive, low-cost and easy to obtain parameters; namely gender, age, body mass index, waist-to-height ratio, systolic and diastolic blood pressures.

The existing research in this field is promising because it shows that a reliable classifier can be designed using easily obtainable information that can be entered on a smartphone, with the exception of blood pressure. Including other MetS risk factors may improve accuracy, however there are currently no studies indicating any success in measuring these parameters using a cellphone application. It will therefore be necessary to investigate more techniques of screening for MetS or its risk factors. The following paragraphs in this section contain several different MetS screening techniques that were investigated by looking at some of the current research into those areas and their relation to MetS.

Diet and Metabolic Syndrome

Kaur showed the protective effect of a vegetarian diet for MetS [4]. This is further shown by Kim [16] in a meta-analysis of observational studies used to assess the association between various meat consumption and the risk of MetS. This meta-analysis included 16 studies and 76,111 participants of which 19,579 had MetS. The study showed that total meat consumption, red meat consumption and processed meat consumption were all positively associated with MetS [16]. This is further supported by Wagh [7], showing that a high fiber, low saturated fat diet with increased daily exercise can reduce the incidence of diabetes by nearly 60%.

Pitsavos [17] looked at a number of different studies on the association between diet, exercise and MetS. In terms of the connection found between dietary patterns and MetS, it was found that a Mediterranean type of diet, high in fruits, vegetables, whole grains, legumes, some fish and low in saturated fats showed lower rates of MetS or associated conditions [17].

Considering the effect that diet has on MetS, this could be considered as both a screening technique and part of the treatment. A patient's current dietary patterns could be monitored either by getting the information in the form of a questionnaire or possibly having the patient use a *"calorie counter"* application similar to *MyFitnessPal* or *Cronometer* if

[10] Fuzzy Neural Network

more detailed information is required [18], [19]. However, a large stumbling block with this type of solution would be obtaining training data. Calorie counting is also an incredibly tedious task, that few people have the patience or dedication to maintain. Requiring this as a necessary step in a screening process would limit use of the application severely.

Exercise and Metabolic Syndrome

As mentioned previously, [17] looked at the association between diet, exercise and MetS. It was shown that the more sedentary a patient was the higher their risk for MetS and that any form of physical exercise could be of benefit [17]. This is further strengthened by findings made by [7] that Diabetes rates are decreased by nearly 60% when improving diet and exercise. Just as in the case of diet, monitoring exercise could play an important role in both screening and treatment of MetS.

Exercise may also be used as part of the screening process, as Sadeghi [20] showed that there is a considerable difference between the maximum HR[11] during exercise between subjects with MetS and subjects without MetS. The maximum HR was also shown to directly correlate wih serum triglyceride levels while inversely correlating with age [20], which means that this may be a valuable marker to be used to model serum triglyceride levels in a non-invasive manner.

In a mobile application, this can be measured in a similar fashion to the mobile application *Cardiio*, developed by researchers at MIT[12], which measures heart rate using a cellphone flash light and camera [21]. The application also gives users a simple 7 minute workout, which can be completed before measuring heart rate again. This would not be advised as a necessary step though, as it would increase difficulty of use, which could lower the odds of the application being used.

Heart rate and Metabolic Syndrome

Maximum HR was found to be related to serum triglyceride levels and MetS [20]. Resting HR was also found to correlate strongly with MetS in seemingly healthy patients in the study by Rogowski [22], suggesting HR can be used when screening for MetS. This correlation was also found by Rana [23] in patients who already had Diabetes and Coronary Artery Disease.

This correlation found between heart rate and MetS suggests that heart rate may be a valuable measurement to be taken for the screening of MetS. This is also a convenient, non-invasive measurement that can easily be taken using a smartphone, as will be discussed in more depth in Section 1.2.2.

[11] Heart Rate
[12] Massachusetts Institute of Technology

Body fat and Metabolic Syndrome

It is known that obesity and specifically abdominal fat plays a major role in MetS [4], [24]. Classifying obesity by BMI[13] alone, as is often done, may not yield the most accurate results, since BMI does not take lean muscle mass into account. Phillips [24] showed that including body fat percentage together with BMI may yield more accurate results than just BMI alone.

Determining body fat on a smartphone could theoretically be done on a visual basis, by creating a classifier that could identify different BF% ranges from a shirtless image. However this tactic may not be necessary, as including waist circumference or weight-to-height ratio may also add the needed information. Since all users may not be comfortable taking pictures of themselves while wearing minimal clothing, this option should be avoided unless absolutely necessary.

1.2.2 Photoplethysmography

As mentioned in Section 1.2.1, MIT developed a mobile application, known as *Cardiio*, that can measure heart rate using the camera and flash light of a smartphone [21]. This is done using a concept known as PPG[14]. PPG is an optically obtained measurement of the volume within an organ or part of the body. A common example of clinical PPG is a pulse oximeter, which emits red and infra-red light onto the fingertip and then measures changes in light absorption. These changes are generally caused by the change in blood volume and oxygen saturation in the blood [25].

Similar to a pulse oximeter a smartphone can be used to measure heart rate and possibly other information by using the flash light as a light source and the camera as the photo detector, as shown in Figure 1.1. This differs from a pulse oximeter in two ways. Firstly a pulse oximeter makes use of infrared light that penetrates deeper and allows for a more accurate reading [25], [26]. Since most smartphones are not able to emit infrared light, another selection within the the visible light spectrum should be made. The second key difference is that a pulse oximeter reads the light passing through the finger whereas a smartphone would only be able to measure reflected light, which will result in an inverted waveform [26]. This can easily be solved in preprocessing.

Heart rate measured using photoplethysmography may be a beneficial tool to be used for screening for MetS. For a more accurate result it may be necessary to determine more of the characteristic conditions, such as blood pressure, triglycerides, and blood sugar. Ivanović [15] conducted a promising study that identified MetS from gender, age, BMI, waist-to-height ratio, systolic and diastolic blood pressures, which suggests that blood pressure may be the only metric that has not been shown to be measurable using a smart phone or known to the user.

Some promise has been shown estimating blood pressure using PPG signals with ANNs

[13] Body Mass Index
[14] Photoplethysmography

Figure 1.1: Example of smartphone based PPG measurement

and other regression techniques [27]–[29]. The studies that looked at this problem found that PPG signals may be an easy continuous and non-invasive way of determining or monitoring blood pressure [27]. But all these studies were performed on information obtained from pulse oximeters and results may not directly transfer to smartphone PPG measurements.

There have been some attempts to determine blood pressure using only a smartphone, as was done by Matsumura, *et al.* [30]. However, by not using the more complex regression techniques used by [27]–[29] their solution lacked the performance found in some other research.

1.2.3 Classification techniques

Promise has been shown in identifying MetS using many different types of classifiers. Barakat suggested a solution where MetS was not only identified using one classifier, but rather a combination of classifiers [11]. This should be kept in mind when deciding which classifier to use. Separate classifiers may also be used for different tasks, for example one classifier may be used to determine blood pressure and another for MetS classification.

Although there are many other possibilities, because of their versatility and the promise that has already been shown in many other studies [15], [27], [28], primarily ANNs will be explored for this dissertation. Different types of neural networks will be considered for different tasks. Given the promise shown by Xing [27] for an FFT[15] based ANN to determine blood pressure and the promise showed by Ivanović [15] in the use of ANNs to identify MetS, this would be an adequate route to follow for this dissertation.

[15]Fast Fourier Transform

1.2.4 What is a screening test?

In order to determine whether an application is a viable screening solution, we first need to define what is meant by screening and what would constitute successful screening.

Herman and Gill [31] define the ideal medical screening test as one that is accessible, simple to use, inexpensive, and associated with minimal discomfort to the person/s being screened. However, they did not provide an ideal or threshold accuracy to be reached. Looking at this criteria in terms of accessibility, smartphones have been a major topic of study for medical devices in recent years due to their popularity. From all the research in this area, a new field known as mHealth has been formed [32]. Their accessibility is the driving factor behind creating smartphone based screening solutions. In order to keep the application as inexpensive as possible, and thereby also increasing the accessibility, a requirement will be to exclude any external sensors. The application will only make use of features and sensors that would be available on a standard smartphone. Ease of use and minimal discomfort will be key design decisions going forward.

In general, the accuracy of screening tests are determined by sensitivity and specificity [33], [34]. In principle, a perfect screening solution should have 100% sensitivity and 100% specificity, meaning it would never make a classification error. But as highlighted by Maxim [33], in practice that may not be the case and the "Gold Standard" is regarded as the best test under "reasonable conditions". It should also be noted that MetS is not a disease or a disorder, but a syndrome, meaning there is no single physical trait or absolute measurement that can be used as the final word. Rather, it is diagnosed based on a group of traits or risk factors that increases the risk of a group of diseases or conditions. The different diagnostic tools also tend to differ slightly, as they may have been formed based on different datasets.

Therefore, the main requirement for the application to be a successful screening solution is to provide a result accurate and detailed enough that the user would have a better understanding of their current health status with regards to MetS and some of its risk factors, which they could use to assess whether a diagnosis from a medical professional would be necessary.

1.3 Research question

With the background in mind and from the information in the literature review, the following question arises:

"How accurately could a smartphone screen for metabolic syndrome, and would this be a feasible solution?"

From this question the following points are derived:

- Can a smartphone be used as a plethysmograph in order to accurately determine heart rate and blood pressure?

- Is it necessary to determine all the characteristic risk factors for MetS (abdominal obesity, fasting blood sugars, serum triglyceride levels and blood pressure) in order to diagnose metabolic syndrome?
- Can information from lifestyle factors, such as diet, exercise or smoking, be used to improve the accuracy of MetS predictions?
- Would an application of this kind be easy enough to use for it to be a feasible solution?

1.4 Methodology

The main question in the problem statement will be addressed by answering the questions that flowed from it. These questions were each posed to investigate each area of the problem. The questions aim to guide the development of the best possible solution.

1.4.1 Can a smartphone be used as a plethysmograph in order to accurately determine heart rate and blood pressure?

Promise has been shown for the use of PPG in determining heart rate and blood pressure. It has also been shown that in some cases a smartphone camera can be used as a photoplethysmograph. As the measurement of heart rate and and blood pressure differ substantially the two topics are discussed separately.

Heart rate

The measurement of heart rate, using a smartphone has already been explored to a large extent. If the results achieved by MIT's *Cardiio* [21] could be replicated, that would be sufficient for our use. The methods used by *Cardiio* are unknown but there are many other resources with different solutions that can be explored [26], [35].

The steps required to determine heart rate are as follows:

1. Capture a video of the user's finger with the flash on.
2. Isolate the appropriate colour (usually green or red) [26].
3. Determine the signal intensity of the PPG waveform by using the average value for each frame or counting a the number of pixels above a threshold value.
4. Determine the fundamental frequency of the PPG waveform.

Determining the fundamental frequency of the PPG is the most important, and most difficult step. There are different ways to do this, with two possible solutions being:

1. Measuring the time between two peaks of the PPG signal to determine the period, and using this period to determine the heart rate.

2. Performing an FFT of the PPG waveform and determining the peak of the FFT.

Figure 1.2 shows the visualisation of a PPG waveform as well as its FFT. These figures illustrate the downsides of both of the techniques that could be used to determine heart rate. Measuring the time between two peaks will be difficult due to the shape of the waveform, as it will be difficult to determine the exact position of a peak. The peak of the FFT on the other hand will be much easier to determine, but may not be accurate enough if the measurement time is not large enough, as this directly influences the resolution of the FFT waveform. A possible solution to this would be to use quadratic interpolation to determine a more accurate peak.

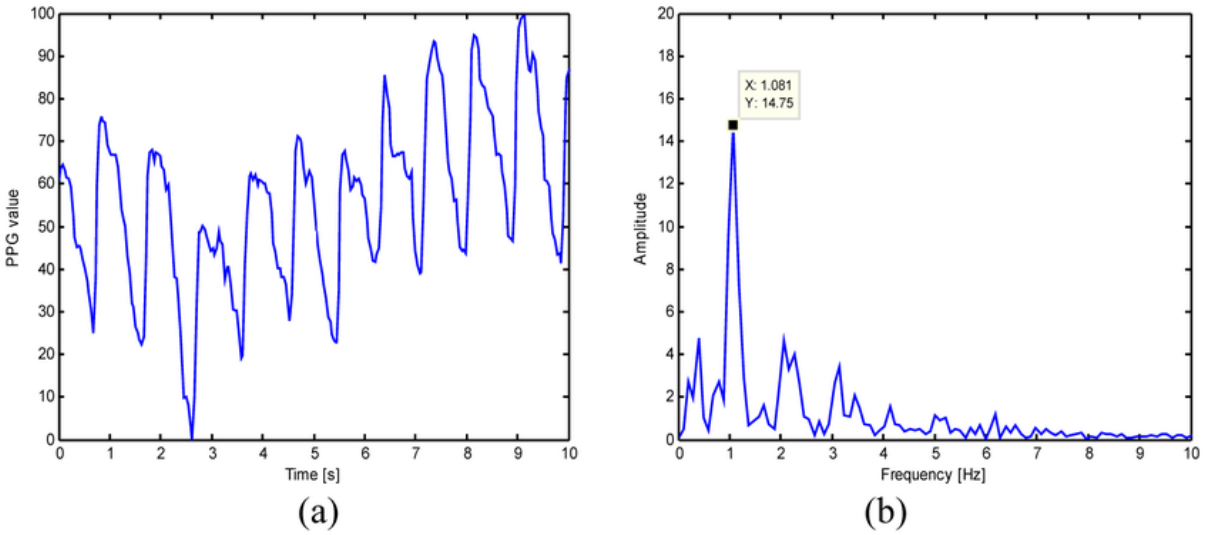


Figure 1.2: Visualisation of (a) PPG waveform measured using a smartphone and (b) FFT of the waveform in (a)

Blood Pressure

Several studies have seen that PPG signals may be a feasible non-invasive way of continuously monitoring BP[16] using ANNs [27], [28]. Keep in mind these studies were performed on pulse oximeter data and measurements using a smartphone may yield different results. However Poh and Chan showed that the same properties may be translated from blood oximeter measurements to smartphone measurements, as the Convolutional Neural Network that was used for the *Cardio Rythm* technology to detect Atrial Fibrillation was trained using down-sampled data from several online databases (including the MIMIC[17] III Waveform Database used by Xing) and this system performed well when using smartphone data as input [27], [36], [37].

[16]Blood Pressure
[17]Multiparameter Intelligent Monitoring in Intensive Care

Assuming that down-sampling database waveforms as was done by Poh [36] will be sufficient, a similar method to Xing [27] with some inspiration drawn from Mousavi [29] will be used used to determine blood pressure. As it is not feasible to create a large enough training set personally, training data would come from an online database, such as the MIMIC III Waveform Database [38], [39]. The waveforms from the database will be normalized and down-sampled to 30 Hz to match the 30 fps frame rate of a typical smartphone camera. For each waveform the FFT will be determined, and the magnitude response, phase response and time domain waveform will be used to train an ANN to determine BP.

Preliminary testing of the ANN was done using a similar database for input data. Real world testing using a cell phone to obtain input data was done to verify the results.

1.4.2 Is it necessary to determine all the characteristic risk factors for MetS in order to diagnose metabolic syndrome?

As stated in the Section 1.1, MetS has five main characteristic risk factors, namely abdominal obesity, high fasting blood sugar, high serum triglyceride levels, low HDL cholesterol and high blood pressure.

When developing a smartphone application that could easily be used, it is important that all the information from the user should be easily obtainable. This may pose a problem with some of the details needed to determine MetS traditionally. Triglyceride and cholesterol levels require blood tests that needs to be done by trained personnel and may be cost prohibitive. Blood pressure and blood sugar can both be measured at home, but require specialized equipment that not many people own. Abdominal obesity should be easier to measure as it is usually determined by elements that people are mostly aware of and could be measured using a scale and a measuring tape, which are both common household items. Each of these conditions will be discussed in further detail to determine whether they could be determined in an inexpensive and non-invasive way.

Abdominal obesity

Abdominal obesity plays a major role in determining MetS and can be determined in different ways, such as BMI, waist circumference, hip-to-waist ratio, and body fat percentage. Each of these have advantages and disadvantages.

BMI is one of the most common methods of determining obesity. It is the technique used by the WHO for their definitions of MetS [6]. BMI is determined by dividing a person's weight by the square of their length. This is good as it takes length into account. The problem is that it does not account for muscle mass, which could greatly influence a person's weight.

Waist circumference is used in the EGIR[18], IDF and NCEP ATPIII definitions of MetS

[18]European Group for the study of Insulin Resistance

and has an advantage over BMI, as it is more independent of muscle mass [6]. The problem with waist circumference is that it does not take skeletal size into account and many individuals may have a larger frame, giving them a larger waist circumference. This problem could be largely addressed by using a waist-to-hip or waist-to-height ratio. Body fat percentage could be one of the most accurate ways of determining obesity [24], but as discussed previously, any benefit that could be seen from building a dedicated fat% classifier would not be enough to justify the added complexity.

The best approach would likely be to use a combination of BMI and waist-to-height ratio, similar to the technique given by the WHO to determine MetS [14].

Fasting blood sugar

Blood glucose levels are usually measured by piercing the finger and taking the measurement from a small blood sample. The meters that are used to take these measurements are usually relatively inexpensive, and there are also many available that can connect with a smartphone. However, these testers are generally not owned by someone unless if they have already been diagnosed with diabetes or another blood sugar irregularity. The pain from this test also makes many users weary to test their blood sugar very often [40]. Some non-invasive methods of measuring blood sugar have been developed, like DiaMonTech that emits light onto the finger and then measures the change in temperature caused by glucose on the skin reacting to the light [41]. Another example is GlucoWise^{TM}, which uses high-frequency radio waves to to measure blood glucose concentration at the capillary level, or GlucoTrack which uses ultrasonic, electromagnetic and thermal sensors to determine blood glucose [40], [42]. These technologies are not readily available yet and the price for these devices are unknown to the author.

So far there is no feasible way known to measure fasting blood sugar in an easy, non-invasive way, using a smartphone. For that reason, a method of determining MetS without the need to measure blood glucose will be used.

Serum triglyceride levels

Serum triglyceride levels are traditionally measured by doing blood tests. There have been several experimental non-invasive measurement techniques that have showed some promise, but none of these will be implementable on a smartphone [43], [44]. Maximum heart rate was shown to directly correlate with serum triglyceride levels by [20]. However, this may not be a feasible metric to use in a mobile screening solution, so only resting HR will be explored. No indication has been seen that this trend also applies to resting HR. It may still be worth including, however, since resting HR was found to correlate strongly with MetS in the study by Rogowski [22]. This suggests that heart rate may be a valuable added feature in the absence of features like serum triglyceride levels, when training the MetS classifier.

HDL cholesterol

Similar to triglycerides, cholesterol is traditionally measured with a blood test. There has been some research into measuring cholesterol using non-invasive methods, such as the PreVu non-invasive cholesterol test which is used to predict skin cholesterol levels [45]. Skin cholesterol, however, does not correlate with serum cholesterol, which makes it inappropriate for detecting MetS. Another non-invasive technique used is to measure cholesterol is bioelectrical impedance. This technique only measures total cholesterol, and even so has been shown to be very inaccurate [45]. This suggests that there may not be a feasible way to determine HDL cholesterol in an easy, non-invasive way.

Blood pressure

Some promise has been shown estimating BP using PPG signals and ANNs [27], [28]. This will be the discussed in detail in later sections of this dissertation where an ANN algorithm will be developed to estimate SBP[19] and DBP[20] using smartphone based PPG measurements.

Proposed solution

So far we have identified that the only parameters that could realistically be used for determining MetS are abdominal obesity, blood pressure and heart rate - which is also a possible indicator of triglyceride levels. This may however still be feasible as shown by [15], meaning all the characteristic conditions may not be necessary for identifying MetS, and that two or three of the conditions could be used together with other information from the user to screen for MetS.

In order to test this hypothesis, an ANN will be designed using TensorFlow in Python and trained using data from the SAPBA dataset collected by researchers from the North West University [46].

Determining results

For all MetS models the results will be analysed with several different performance indicators, namely accuracy, precision, recall and F1 score. The equations used to calculate these scores are give in equations 1.1 to 1.4.

[19] Systolic Blood Pressure
[20] Diastolic Blood Pressure

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{1.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{1.3}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1.4}$$

where TP is the number of true positive predictions, TN is the number of true negatives, and FP and FN are the number of false positives and false negatives respectively. As accuracy is the score most often used, and most intuitive to understand, it will be the main performance indicator, with the others used to recognise biases in the data or the model.

In most medical applications, screening tests are assessed based on sensitivity and specificity. Sensitivity is mathematically equivalent to recall. Specificity, like precision, also looks at the false positive rate but in relation to true negatives. Therefore, sensitivity and specificity can be determined using the following equations:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{1.6}$$

The first set of variables are used in order to initially assess the performance of the model. Precision and recall are more directly comparable to each other, meaning, together with the F1 score will more clearly indicate most biases in the data. This method only becomes less effective in cases with very few true negatives. When assessing the model in relation to other diagnostic tools, however, sensitivity and specificity are used.

1.4.3 Can information from lifestyle factors be used to improve the accuracy of MetS predictions?

The study by Kaur [4] assessed age, gender, social status, employment, education, family history, physical activity, dietary habits, alcohol, sleep, body mass index and stress showed correlation between most of these areas and MetS. This shows that there is definitely a connection between these lifestyle habits and MetS, and implies that it may be possible to predict MetS when considering these lifestyle factors. In a study by Ushida [12] an FNN was used to find that alcohol and cigarette use highly correlated with the most significant risk factors for the development of MetS.

As seen in the study by Jahandideh [47], it may be possible to evaluate non-invasive measurements or their effects from lifestyle factors, implying that lifestyle factors, such as drinking, smoking or exercise could be used as predictors for MetS.

In practice these factors could be determined using a questionnaire on an application, where the user would fill in information like age, sex, height, weight, waist circumference, smoking status and answer questions based on dietary and exercise habits. The information from this questionnaire would then be used as input variables to the final classifier, together with other data retrieved from measurements such as PPG.

Figure 1.3 shows a flow diagram of the point of view from which the user should experience the software. The user would enter his/her personal health data, take the necessary PPG measurements and then view the results that the software has determined.

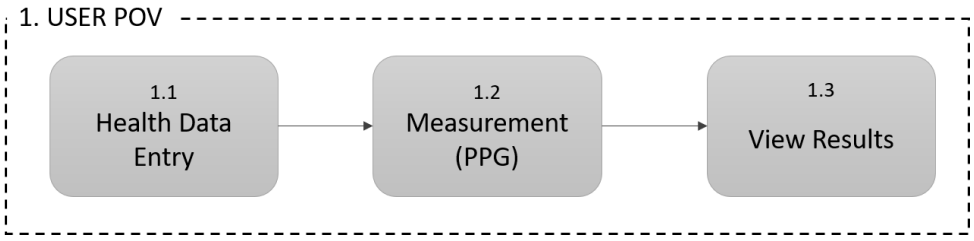


Figure 1.3: Order of operations from the user's point of view

The process that the software uses to determine whether the user is at risk for MetS can be seen in Figure 1.4. The software makes use of the information supplied by the user in Figure 1.3, and is referenced using numbers (1.1, 1.2), to determine the output (1.3). The software determines heart rate and blood pressure which are used with the remaining health data received from the questionnaire to determine the MetS result.

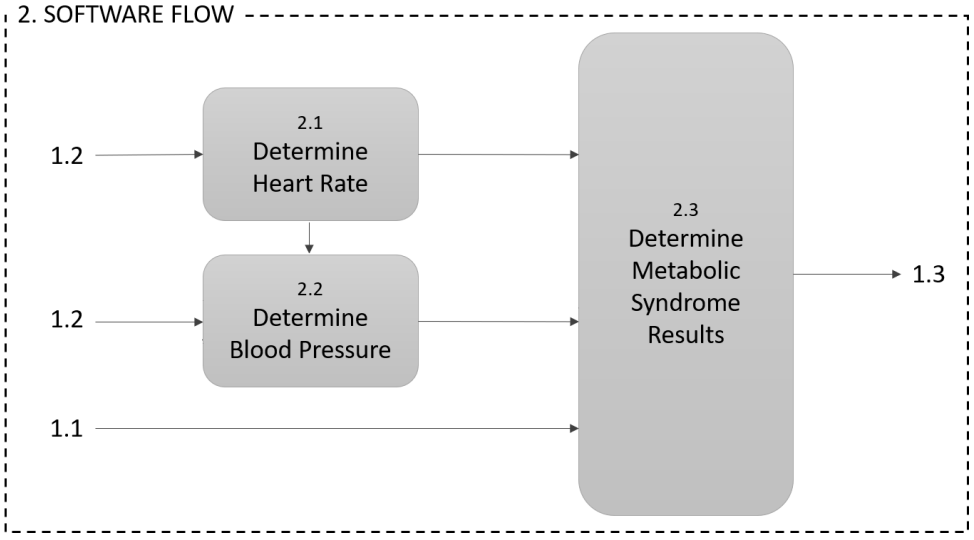


Figure 1.4: Flow diagram of operations to be completed by the software

As with the previous classifier, this model was trained and tested using data obtained from the SABPA dataset [46]. Tensorflow is used for the machine learning algorithms,

and models are converted to TensorFlow Lite for its effective integration with Android Studio. These models are built using Python in the Jupyter Notebook IDE, due to its live code, visualisation and markdown text features that make it ideal for experimentation and rapid code adjustments.

1.4.4 Would an application of this kind be easy enough to use for it to be a feasible solution?

In order to overcome the need to rely on simple intuition to determine usability a number of observable and quantifiable metrics need to be determined.The ISO 9241-11 standard defines usability as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* [48]. And thus the usability of the application will be measured in terms of effectiveness, efficiency and satisfaction.

Effectiveness

The effectiveness with which users are able to use an application includes measures for completion rates and errors, where completion rate is often referred to as the fundamental usability metric [49]. Completion rate is calculated as a ratio between the number of tasks completed successfully and the total number of tasks undertaken, as seen in the following equation [49].

$$Effectiveness = \frac{Number\ of\ tasks\ completed\ successfully}{Total\ number\ of\ tasks\ undertaken} \times 100\% \tag{1.7}$$

Although one should always aim for a completion rate as high as possible, according to a study carried out by Sauro [49], the average task completion rate is 78%. However, it was also observed that the completion rate is highly dependent on the context of the task being evaluated.

In order to test effectiveness when conducting the tests of the final software, users will not be given any instructions, other than those received by the mobile application. In the case that the user cannot complete a task he/she will ask the facilitator for help and that task will be marked as unsuccessful. At the end of the test the effectiveness will then be determined and an average effectiveness could then be determined between all users.

Efficiency

According to ISO-9241, product Efficiency is defined as "resources spent by user in order to ensure accurate and complete achievement of the goals" [48]. In this case the resource used by the user is time. Time based efficiency is determined as follows:

$$\overline{P}_t = \frac{1}{NR} \sum_{j=1}^{R} \sum_{i=1}^{N} \frac{n_{ij}}{t_{ij}} \tag{1.8}$$

where N is the total number of tasks, R is the number of users, n_{ij} is the result of task i by user j ($n_{ij} = 1$ for successful task and 0 for unsuccessful) and t_{ij} is the time spent completing the task. The result has a unit of tasks/sec, which does not give much meaning, unless being compared to a similar system. Relative efficiency may be calculated by also using the time of the tasks that were not completed.

$$\overline{P} = \frac{\sum_{j=1}^{R} \sum_{i=1}^{N} n_{ij} t_{ij}}{\sum_{j=1}^{R} \sum_{i=1}^{N} t_{ij}} \times 100\% \tag{1.9}$$

This results in a relative value that could be compared to other cases, and could therefore be used to draw a conclusion.

Satisfaction

Measuring user satisfaction is done using a standardized satisfaction questionnaire. The questionnaire will be filled in at the end of the test, to save time and as that should be sufficient to get each user's overall opinion of the application. A SUS[21] is used to question the users, as this is a simple test that has been shown to accurately represent the opinion of users [50]. The results from the questionnaire are then compared to the SUS in Figure 1.5, which was created based on information by [50], to draw a conclusion on the satisfaction of the user.

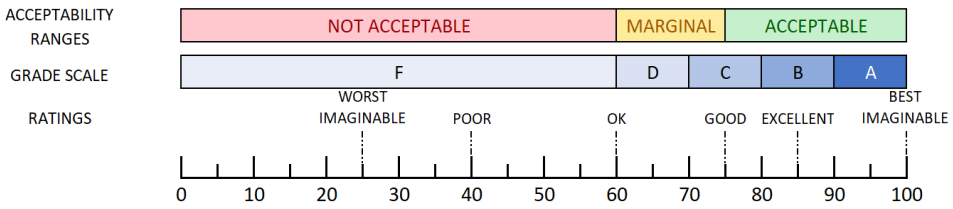


Figure 1.5: Grade rankings of SUS scores

1.5 Chapter overview

The chapters in the dissertation are discussed in this section.

[21] System Usability Scale

1.5.1 Chapter 2: Photoplethysmogrophy measurements

This chapter describes the process of PPG measurement on a smartphone and the techniques used to determine heart rate and blood pressure from such measurements. Two methods of measuring heart rate are investigated. A hydraulic analogy model is developed to express the relationship between blood pressure and PPG in terms of an electronic circuit, which is used to guide the feature selection process. An ANN is used to determine blood pressure from the features extracted from the PPG waveform.

1.5.2 Chapter 3: Metabolic Syndrome Classification

This chapter discusses the design of the ANN to be used to classify MetS status. In order to answer the questions posed in the problem statement, two separate models are built. The first model uses primarily quantitative inputs from features that are ordinarily used to determine MetS, excluding those features that cannot be measured using a smartphone or wouldn't be ordinarily known by the user. The second model includes lifestyle factors, such as activity level and smoking or drinking habits, with the features used in the first model.

1.5.3 Chapter 4: Experimental procedure

In this chapter the experimental procedure used to validate the software designed in chapter 3 will be discussed. The questions in the problem statement will largely influence this chapter, as the results from the experiments should provide answers to these questions.

1.5.4 Chapter 5: Results

The results from the experiments in chapter 4 are documented in this chapter. The results will be divided according to heart rate, blood pressure, MetS and usability.

1.5.5 Chapter 6: Conclusion

The results are discussed and reference will be made back to the problem statement in order to determine the successfulness of the study.

CHAPTER

2

PHOTOPLETHYSMOGRAPHY MEASUREMENTS

This chapter describes the process of PPG measurement on a smartphone, as well as the techniques to determine heart rate and blood pressure from such measurements. Two methods of measuring heart rate are investigated. A hydraulic analogy model is developed to express the relationship between blood pressure and PPG in terms of an electronic circuit, which is used to guide the feature selection process. Finally, an ANN is used to determine blood pressure from the features extracted from the PPG waveform.

This dissertation assumes a working knowledge of artificial neural networks (ANNs). For a reasonable understanding of neural networks, consult the first chapter of the book, *Neural Networks for Identification, Prediction and Control*, by Duc Pham and Xing Liu [51]. Another good resource is the video series on neural networks by Grant Sanderson on the YouTube channel, 3Blue1Brown [52].

2.1 Relevant literature

2.1.1 Heart rate

As discussed in the introductory chapter, PPG is a simple optical technique used to detect blood volume changes in the microvascular bed of tissue without the need for any invasive procedures [53]. Researchers from MIT developed a mobile application, known as

Cardiio, that can measure heart rate using the camera and flash light of a smartphone to do PPG measurements [21]. A more common clinical example of PPG is a pulse oximeter, which emits red and infra-red light onto the fingertip and then measures changes in light absorption caused by the change in blood volume and oxygen saturation in the blood [25].

Other than *Cardiio*, there are many other smartphone applications using a camera to measure a PPG signal for heart rate estimation. A meta-analysis by De Ridder [35] looked at 14 studies that tested the accuracy of different smartphone based HR monitoring applications. The accuracy in all cases were always above 90% with an average accuracy of 95%.

2.1.2 Blood Pressure

Some promise has been shown estimating blood pressure using PPG signals and ANNs [27]–[29]. The studies that looked at this problem found that PPG signals may be a simple non-invasive way of continuously measuring or monitoring blood pressure. However, the majority of these studies were only performed on pulse oximeter data and it is unknown if a smartphone would achieve the same results.

One popular method, used by Tabei, *et al.* [54], is to use pulse transit time between two seperate PPG signals to estimate BP. A similar method was used by Liu, *et al.* [55], with the addition of PCA[1] to detect motion artefacts in the signal. While both achieved high accuracy with their models, a major downside is that they required two separate, external PPG sensors.

In recent years there have also been some attempts to determine blood pressure using only a smartphone camera, as was done by Matsumura, *et al.* [30]. They obtained an accuracy of more that 70 % using only a linear regression formula. Their solution was based on a formula, well known in cardiology and very intuitive from an engineering perspective, which will be explored later when we develop a hydraulic analogy formula in Section 2.4.1. The formula [30] used was:

$$BP = CO \times TPR \tag{2.1}$$

where CO is cardiac output, which can be thought of blood flow rate, and TPR[2], also known as systemic vascular resistance, which is the combination of resistance to flow from arteries, veins and other peripheral organs or tissue. This can be turned into a linear regression formula by using logarithmic rules:

$$ln(BP) = ln(CO) + ln(TPR) \tag{2.2}$$

They then used the knowledge that HR correlates with CO and mNPV[3] correlates with

[1]Principal Component Analysis
[2]Total Periphery Resistance
[3]Modified Normalized Pulse Volume

TPR to create a linear formula of form:

$$BP = (a \times ln(HR)) + (b \times ln(mNPV)) + c \tag{2.3}$$

A major drawback to this approach was that many of the assumptions they made, would rarely be the case in practice. An example of this is that CO is dependant on both HR and stroke volume, which would be vastly different from person to person and possibly even from day to day in the same person. The method also used the same formula with only different coefficients for DBP and SBP, which in many cases may be vastly different. The use of mNPV also meant that per device calibration would be necessary. Despite these problems, much could be learnt from this study, as some of its principles were what inspired the design approach in Section 2.4.1.

While the motivation used for the ANN designs by [27] and [29] often went somewhat outside what could be considered a typical engineering knowledgebase, a lot of value could be used from these solutions and the results they obtained. Xing, *et al.* [27] used an FFT based ANN to obtain SBP and DBP from a single PPG wave. Mousavi, *et al.* [29] wished to expand on the work by Xing and others to make a more robust algorithm, using multiple types of inputs, including a full time domain waveform, filtered using FFT and reduced dimensions using PCA.

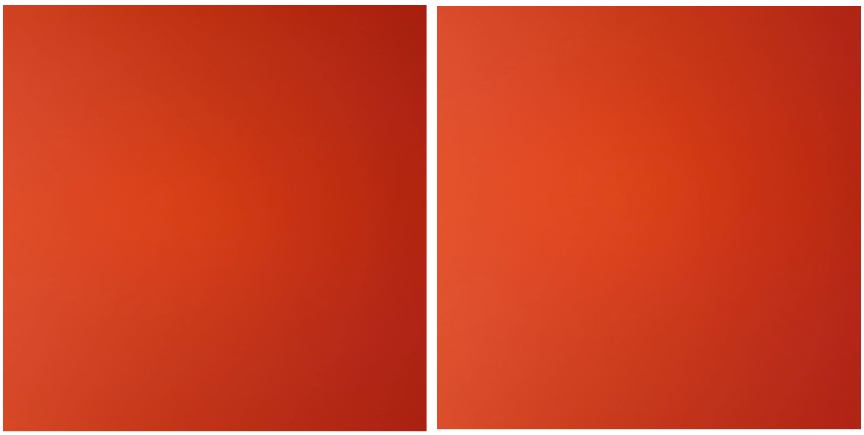
Based on the above-mentioned literature, it was decided to use an ANN to estimate blood pressure from smartphone-based PPG measurements. The model that was decided on would be a combination of previous studies, using FFT and time domain waveforms as input into the ANN.

2.2 PPG on a smartphone

In a healthcare setting the most common PPG measurement device is a blood oximeter, which works by clipping over a patient's fingertip, transmitting red and infrared light on one side of the finger and measuring the light on the opposite side using a photodiode. The reason for using red and infrared light is because absorption of light at these wavelengths differs significantly between blood saturated with oxygen and blood lacking oxygen, so red light can be used as a reference to infrared [56]. Measuring oxygen saturation by transmitting light through a body part is known as transmissive pulse oximetry, but there is another method (although less popular) known as reflectance pulse oximetry. Although, clinically, blood oximeters are mostly used for measuring oxygen saturation in blood, PPG refers specifically to the change in blood volume rather than oxygen saturation within the blood. There are an increasing number of devices on the market, such as fitness watches, that use PPG measurements to determine heart rate.

PPG measurement on a cell phone is done by placing your fingertip over the camera lens, covering the lens as well as the flash light. The flash light acts as a light source and the

camera measures the reflected light intensity. The resulting systolic[4] and diastolic[5] images recorded can be seen in Figure 2.1. When comparing Figures 2.1a and 2.1b by eye, the difference is not evident, however, can sometimes be noticed in a video. With 8 bit pixel values, these changes tend to be easily detected and even more so when looking at all the pixels in a frame.



(a) Systolic point (b) Diastolic point

Figure 2.1: Example images recorded using smartphone camera

Each frame that the camera captures is colour isolated. Either red or green light can be used and is discussed in more detail in Section 2.2.1. To compute the PPG signal we used a threshold for the appropriate colour components for each frame obtained and computed the number of pixels that surpassed the threshold, similar to the method proposed in [57]. The threshold T was established as 95% of the range between the minimum and maximum values during a calibration stage in the first 5 seconds of system operation. That is:

$$T = \overline{max}(I) - \frac{1}{20}(\overline{max}(I) - \overline{min}(I)) \tag{2.4}$$

Figure 2.2a shows a colourmap of the red channel of the image from Figure 2.1, which gives a clearer perspective of the distribution of values from a finger image. Figure 2.2b displays a threshold mask that would be associated with Figure 2.2a, depending on the exact T value that was determined in the calibration stage.

The amplitude of the PPG signal is determined by counting the number of pixels with red values over the threshold value. The final PPG signal will be the inverse of this however, since this is a reflective measurement and conventional PPG signals are based on transmissive signals. The amplitude of this final PPG signal is given by the following formula:

$$PPG_i = -\sum_{n=1}^{N} \delta[n] \quad \text{for } \delta[n] = \begin{cases} 0, & x_n < T \\ 1, & x_n \geq T \end{cases} \tag{2.5}$$

[4]Point in the heart beat cycle when the heart contracts, i.e. peak pressure
[5]Point in the heart beat cycle when the heart rests, i.e. lowest pressure

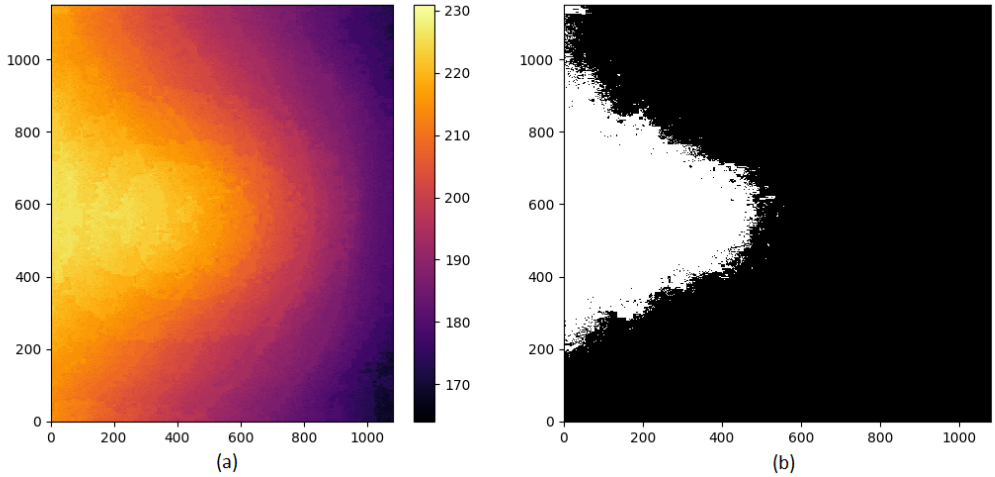


Figure 2.2: (a) Colourmap of red signal from finger image with (b) example threshold mask

where x_n is the colour value of the specific pixel and N is the number of pixels in the frame.

Because the amplitude of the signal is highly dependant on factors like the type of smartphone, the external lighting or skin tone, the signal will be normalised before it is used for further processing. The normalisation is described by the following equation:

$$PPG_{norm} = \frac{PPG_i - PPG_{min}}{PPG_{max} - PPG_{min}} \tag{2.6}$$

2.2.1 Colour selection

There are several factors that determine the optimal light source for a PPG measurement. It is immediately clear from Figure 2.1 that the colour reflected most by the skin is red. Red light also penetrates skin the deepest, which explains why it would be used as a baseline in transmissive pulse oximeters. Moving further up the spectrum to green light, most is absorbed by the skin but it penetrates deep enough and enough is reflected to still obtain a PPG signal. Blue light, however, is almost completely absorbed, barely penetrating the skin and therefore getting minimal signal.

Jonathan and Leahy used a smartphone for pulse rate measurement, and they assessed that the green channel provides a stronger AC signal than the red channel [58], [59]. However, their research involved a limited number of smartphones and further research has shown that the typical color range varies for different models of smartphones. It has been shown by [26] that only the red channel has similar characteristics for different smartphone models while the green and blue may vary dramatically.

Considering that red light penetrates the skin much better than green light, red would be a better option for obtaining the PPG signal in cases where the light source is not

directly next to the camera lens. This is often the case in modern smartphones that have multiple camera lenses. This was also the case on the smartphone used during the testing for this dissertation, which was a Samsung Galaxy A50 that had 3 lenses (as seen in Figure 2.3) with the center lens being the main camera and therefore resulting in a gap of approximately 15 mm between the lens and the flash.

Figure 2.3: Camera lenses for Samsung Galaxy A50

For this reason red was chosen as the colour that would be used to form the PPG signal. Blue and green channels were still used, however, as references to determine whether the user is operating the application correctly, which is described in the following section.

2.2.2 Signal verification

When taking measurements on a smartphone application, only the user is in control of the proper operation and correct measurement process. This is in contrast to a clinical environment where the medical staff can supervise the whole procedure and detect when it goes wrong. The wrong position of the fingertip on the smartphone camera or its absence, finger movement during the measurement or even changing the force applied by the fingertip to the lens may cause incorrect measurement and, as a result, the application gives false alarms or misses a diagnosis. To prevent incorrect health parameter measurement, the application was designed to automatically detect cases of improper usage and provide feedback to the user.

The validation process that was followed is based on the method used by [26]. The following formulas are used throughout the validation process:

$$mean(R) - \sigma_R \geq R_{min} \tag{2.7}$$
$$mean(G) - \sigma_G \geq G_{min} \tag{2.8}$$
$$mean(G) + \sigma_G \leq G_{max} \tag{2.9}$$
$$mean(B) + \sigma_B \leq B_{max} \tag{2.10}$$
$$\sigma_R, \sigma_G, \sigma_B < \sigma_{max} \tag{2.11}$$

where R is the pixel value for red, G is green, B is blue and σ refers to standard deviation across the pixels of a single frame.

The first validation step for each frame would be to determine if the user has correctly placed their finger on the camera lens. During the calibration stage this is done by substituting the standard values, determined experimentally by Kurlylyak [26] for a range of phones, in equations 2.7 to 2.11. These are $R_{min} = 128$, $G_{min} = 10$, $G_{max} = 128$, $B_{max} = 128$ and $\sigma_{max} = 40$. When a user takes a measurement, these values will be calibrated a few seconds into the measurement to the specific device and measurement. Straying from the calibrated values will alert the application of movement or other issues during the measurement, which would cause the application to enter into an error state and alerting the user to restart the measurement.

2.3 Heart rate

Determining heart rate from a PPG wave can be done in several ways, two of which were considered here. The first method is to measure the time between two peaks of the PPG wave and the second is using the frequency at the maximum value in an FFT of the PPG wave.

Using the time based method, heart rate is obtained by dividing 60 by the peak to peak time period (T) of the PPG waveform, as seen in the following formula:

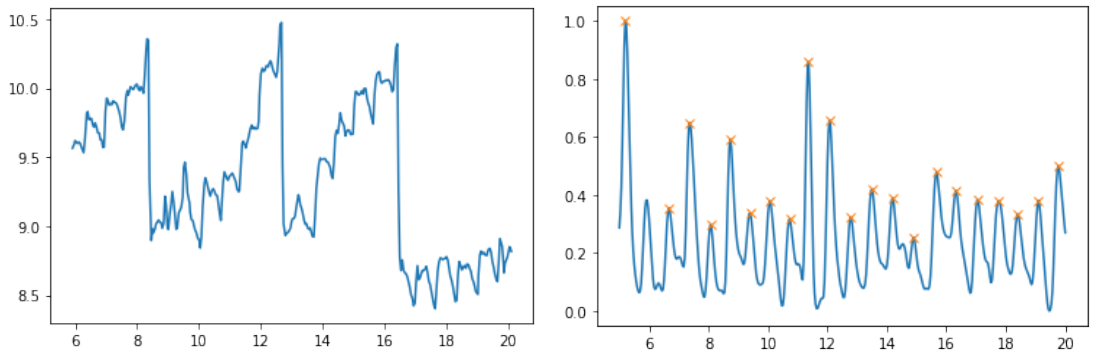
$$HR = \frac{60}{T} \tag{2.12}$$

In order to reduce errors, a second order bandpass filter was also applied to the signal before doing peak detection. Only frequencies between 0.5 and 3 Hz were passed, which corresponds to 30 to 180 BPM. The output formula derived from the Z-transform is as follows:

$$\begin{aligned} y[n] =& 0.04948996x[n] - 0.09897991x[n-2] + 0.04948996x[n-4] \\ &+ 3.16980927y[n-1] - 3.85528971y[n-2] \\ &+ 2.1599809y[n-3] - 0.47759225y[n-4] \end{aligned} \tag{2.13}$$

An example of a PPG signal with a shifting peak amplitude, from low frequency interference, can be seen in Figure 2.4a and the resulting waveform from the bandpass filter in Figure 2.4b.

The position of the peaks were determined by looking for downward zero-crossings in the smoothed first derivative with peak amplitudes that exceed the average signal value over 45 samples (1.5 s). This was done for several periods, then outliers were removed to reduce the odds of missed peaks or incorrectly picking up diastolic notches, and an average T was determined.



(a) PPG signal with low frequency interference

(b) PPG signal after filtering and peak detection

Figure 2.4: Example of PPG signal with low frequency interference with same signal after filtering and peak detection

The second option is to find the peak from an FFT of the PPG and multiply the frequency of that point by 60, as in:

$$HR = 60 \times f_{max(FFT)} \tag{2.14}$$

Something to consider when determining heart rate is the measuring resolution of the method being used. Both are dependent on the sampling frequency (f_s), which in our case will be a constant 30 Hz. The time based method will also be dependent on the number of pulses that can be measured (N_p) and will have the following measurement resolution:

$$\frac{60}{f_s N_p} = \frac{2}{N_p}\text{HR} \tag{2.15}$$

The measuring resolution of the FFT based method also depends on the sampling frequency and is further dependant on the number of samples (N_s). This is due to the definition of FFT, in that the highest frequency is the sampling frequency and there are as many FFT samples as in the input which means that the smallest frequency change is f_s/N_s. This implies that the resolution of an FFT based measurement can be calculated as follows:

$$\frac{60 f_s}{2N_s} = \frac{30}{t_s}\mathrm{HR} \tag{2.16}$$

With this in mind we compare the two methods with an example where a patient has a heart rate of approximately 60 BPM, meaning that each beat would take one second. That would give the time based method a measuring resolution of 2 BPM and the FFT method 30 BPM, which shows that FFT is not suitable for short time intervals. For a 10 s measurement the resolution for the FFT method is 3 BPM which is usable but still much higher than the 0.2 BPM of the time based method. The FFT method's resolution can be improved by using an interpolation algorithm, but this would unnecessarily complicate the process and waste processing time.

Because initial measurements would already be known and wouldn't need to be recalculated on each beat, time based measurements are also more efficient during continuous measurement. As will be seen in the following section (2.4), an FFT algorithm was used in the blood pressure algorithm, therefore it was considered to still use an FFT based method as an extra validation step at the end of the measurement without adding any complexity to the process, however, this method was found to be less reliable so only the time based method was used.

2.4 Determining blood pressure

2.4.1 Hydraulic analogy

In order to fully understand how blood pressure is determined by other elements within the circulatory system, the main components were represented in terms of their hydraulic and electrical counterparts. This helped to guide the development of the blood pressure estimation algorithm that was used.

The main elements of the circulatory system are the heart, blood vessels (veins, arteries, venules, arterioles, capillaries), muscle tissue and organs. Each time the heart beats, blood is delivered from the heart, through the arteries to arterioles to capillaries, which provides blood to the different tissue and organs, from which blood flows via venules to veins, back to the heart.

Table 2.1: Analogous electrical and hydraulic equations

Type	**Hydraulic**	**Electrical**
Quantity	Volume (V: m^3)	Charge (q: C)
Quantity flux	Volumetric flow rate (Q: m^3/s)	Current (I: A)
Potential	Pressure (p: Pa)	Potential (ϕ: V)
Linear model	Poiseuille's law	Ohm's law

If considering the circulatory system in terms of a hydraulic system, parallels can be drawn to different engineering realms, such as electrical circuits as can be seen in Table

2.1. The heart is a pump that provides pressure, thereby causing blood to flow through the different blood vessels to organs and tissue and then back to the heart. The different vessels and paths through which the blood flows have different diameters and lengths, which impede the flow.

The equivalent of Ohm's law ($V = IR$) for hydraulics is the Hagen–Poiseuille law, which states:

$$\Delta p = \frac{8\pi\mu LQ}{A^2} \tag{2.17}$$

Where Δp is pressure difference (equivalent to voltage), Q is volumetric flow rate (equivalent to current), μ is viscosity, L is the length of the pipe (or blood vessel) and A is the cross sectional area of the pipe, which means the flow resistance will be:

$$R_{hydro} = \frac{8\pi\mu L}{A^2} \tag{2.18}$$

Compare this to electrical resistance, which is calculated by:

$$R = \frac{\rho L}{A} \tag{2.19}$$

Where ρ is resistivity of the material. This shows an area where the analogies may seem to diverge, since the hydraulic equivalent to resistivity would be $(8\pi\mu)/A$, meaning hydraulic resistance varies more with cross section than electrical resistance. This does not, however, affect our application. All that should be noted in this case is the dependence on length, cross section, and viscosity.

If we represent the human circulatory system as a simple circuit where the heart is a voltage source and the systemic vascular resistance is an electrical resistor, then Ohm's law can be used to determine the pressure provided by the heart if we know the flow rate and vascular resistance. This achieves a formula equivalent to that was used by Matsumura (equation 2.1) when determining blood pressure [30]. This way blood pressure can be determined using the following formula:

$$MAP = Q_C \times TPR \tag{2.20}$$

Where MAP is the mean arterial pressure, Q_C is the average cardiac output (measured in l/min) and TPR is the total resistance to flow, as discussed in Section 2.1.

While this model may be adequate in representing the problem as an average over a period of time, it becomes more complex when assessed in more detail. This is because TPR and Q_C don't remain constant over time. An increase in cardiac output would result in a greater arterial pressure. Since blood vessel walls are elastic, an increased pressure would cause the cross section to grow. As seen earlier, a larger cross section would result in a decreased resistance, which would then result in a decreased pressure. Treating the

arterial wall like a flexible diaphragm means that it can be represented by a capacitor. This results in a low-pass filter circuit, as shown in Figure 2.5a with a simulated output example in 2.5b. See Figure 2.7 for an illustration of actual ABP[6] and PPG signals.

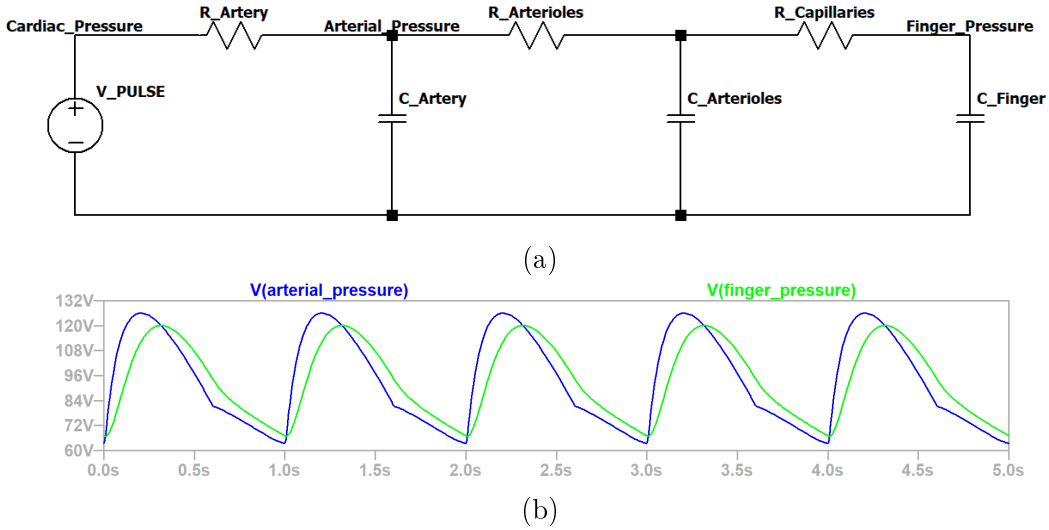


Figure 2.5: (a) Electrical circuit representing circulatory system with (b) simulated outputs for arterial pressure (blue) and peripheral finger pressure (green)

In reality, the human circulatory system would be far more complex with a much higher order filter circuit containing many more parallel branches. Furthermore, in the biological model there is also a causal relationship between the volume change and the resistance with resistance varying with time depending on the volume and pressure. This 3rd order filter, however, will be sufficient in understanding the relationship between blood volume and blood pressure.

In searching for an appropriate input voltage (pressure) to use for the simulation in Figure 2.5, the diagram in Figure 2.6 was found in the Guyton and Hall Textbook of Medical Physiology [60]. The input that was approximated was the aortic pressure (top black dotted line). Figure 2.6 was also used in research [61], which simulated the circulatory system using passive filter circuits. Their simulation was far more in depth with filters going up to the 24th order. While the contents of the article go far beyond the scope of this dissertation, it may be something of interest to readers who wish to know more about such physiological models.

There are several formulas associated with the circuit diagram in Figure 2.5a that help to give clarity to the scope of the problem. The first of these is capacitance:

$$q = CV \tag{2.21}$$

where q is electrical charge, C is capacitance and V is voltage. Using Table 2.1 this formula teaches us that volume and pressure are directly proportional. Equating capacitance and

[6] Arterial Blood Pressure

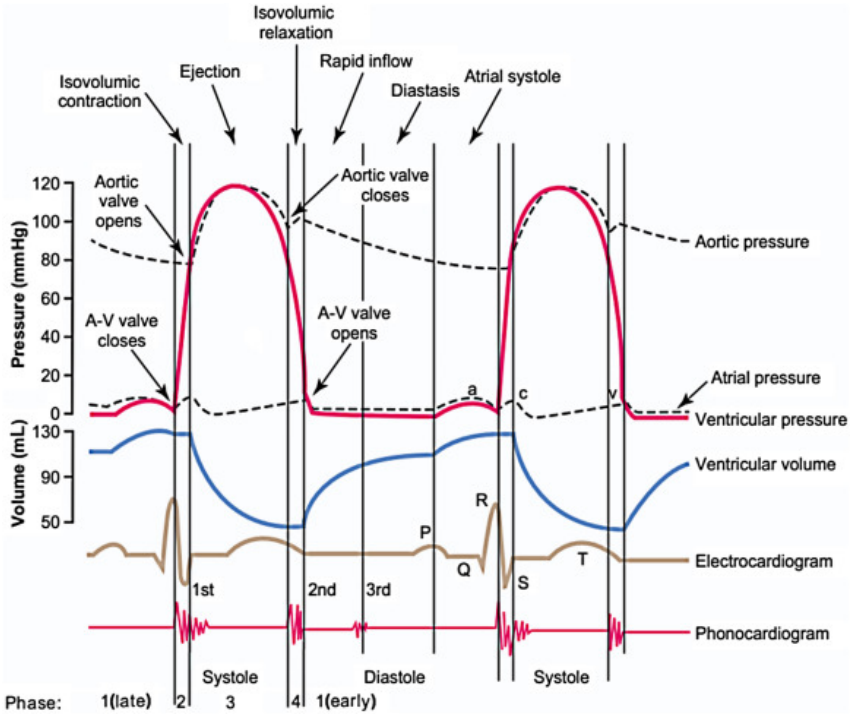


Figure 2.6: Cardiac cycle showing left atrial pressure, left ventricular pressure, aortic pressure ventricular volume, the electrocardiogram, and the phonocardiogram from [60]

elasticity means that vascular elasticity is directly proportional to blood volume and inversely proportional to blood pressure.

Furthermore, considering that this is a low-pass filter, each RC circuit has the following transfer function:

$$\frac{V_{out}}{V_{in}} = \frac{1}{1 + sRC} \tag{2.22}$$

we can therefore determine that the arterial pressure (V_{ABP}) will be:

$$V_{ABP} = \frac{V_{in}}{1 + sR_{artery}C_{artery}} \tag{2.23}$$

and finger blood volume (q_{PPG}) is a function of finger pressure (V_{FP}) so, if capacitance remains constant, will therefore be:

$$q_{PPG} = V_{FP}C_F \tag{2.24}$$

Considering that there is a second order low-pass filter between V_{ABP} and V_{FP}, the following transfer function would apply.

$$\frac{V_{FP}}{V_{ABP}} = \frac{1}{1 + sb_1 + s^2 b_2} \tag{2.25}$$

where:

$$b_1 = \tau_A + \tau_F \tag{2.26}$$
$$b_2 = \tau_A \times \tau_{FA} \tag{2.27}$$

and the RC time constants are:

$$\tau_A = C_A R_A \tag{2.28}$$
$$\tau_F = C_F(R_F + R_A) \tag{2.29}$$
$$\tau_{FA} = C_F R_F \tag{2.30}$$

where C_A and R_A refer to arteriol capacitance and resistance, and C_F and R_F refer to capillary/ finger capacitance and resistance. This results in the following transfer function:

$$\frac{V_{FP}}{V_{ABP}} = \frac{1}{1 + s(R_A C_A + R_A C_F + R_F C_F) + s^2 R_A R_F C_A C_F} \tag{2.31}$$

We can then combine equations 2.24 and 2.31 to obtain the transfer function to get from finger blood volume to arterial blood pressure as:

$$\frac{V_{ABP}}{q_{PPG}} = \frac{1}{C_F}(1 + s(R_A C_A + R_A C_F + R_F C_F) + s^2 R_A R_F C_A C_F) \tag{2.32}$$

Which tells us that the effects of the system will not be limited to changes in amplitude but will be affected by changes in frequency as well, as can be seen from the low-pass filter effect in Figure 2.5b as well. None of the variables required (peripheral resistance or elasticity) in this transfer function can be determined non-invasively so therefore a direct equation will not be possible to determine blood pressure. These variables also differ greatly from person to person and even day to day. However, we can predict that arteriol and capillary elasticity will be much lower due to the surrounding muscle tissue which means we can expect to see the two signals closely correlate at low frequencies, though a time lag is to be expected. And since these variables will have an impact on the frequency response of blood volume, an FFT may provide enough information to derive blood pressure using machine learning without knowing systemic vascular resistance or elasticity.

2.4.2 Data analysis and preprocessing

The dataset that was used for training, validation and initial testing of the blood pressure algorithm is the MIMIC III Waveform Database [39], [62]. The database contains 67,830 record sets for approximately 30,000 ICU patients. Most of the record sets contain a waveform record containing digitised signals which usually include ECG, ABP, PPG, respiration, etc., and often there is also a record containing time series of periodic measurements.

For the purposes of creating the blood pressure algorithm the only signals required are the ABP and PPG signals. ABP in this case is an invasive continuous measurement made using an arterial catheter connected to a pressure transducer. Both of these signals were sampled at 125 Hz, which means training and test data must be resampled to 30 Hz to match a standard smartphone. The PPG was measured using a blood oximeter, meaning that it is a transmissive signal whereas that measured on a smartphone is reflective. This is taken into account in the smartphone measurement by inverting the signal.

Data sifting

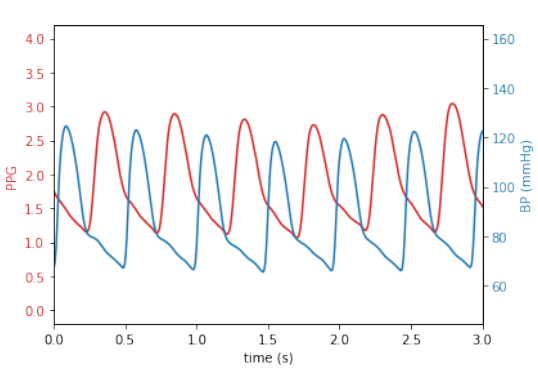
The first step in finding all the relevant datasets was to go through all the patients' data to find those that have both ABP and PPG signals. Each of these signals also had to be tested for quality. Any null values from either the PPG or ABP signals would mean that section must be removed from both waveforms. Signals with too great a proportion of null values were disregarded entirely. This was only one part of the errors that could be found in the dataset. Often patients would move or were treated and these artefacts could be picked up by the measuring devices, causing them to become inaccurate.

To analyse the data more thoroughly all the remaining patients' datasets were split into 10 seconds samples. In this way each sample could be analysed and kept or eliminated based on the signal quality.

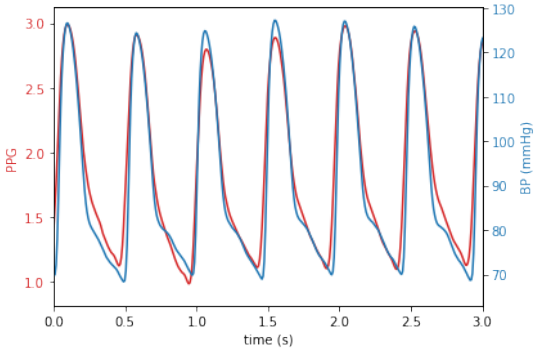
PPG and ABP correlation

As seen in Section 2.4.1 we can expect PPG and BP to correlate well. Figure 2.7a shows this correlation and that PPG lags behind the ABP waveform. This lag is partly due to the lag effect caused by arteriol and capillary elasticity and resistance, as discussed in the previous section, but will differ from the model since fluid propagation time is much slower than electrical current. This lag was removed by using cross correlation between PPG and ABP and determining the exact lag amount by finding the peak correlation with positive lag, as seen in Figure 2.7c. It has to be positive because PPG always lags behind ABP. The first part of the PPG wave up to the lag time is then cut out and both the ABP and PPG signals are trimmed to 8 seconds sample length to accommodate for the maximum possible lag time lost.

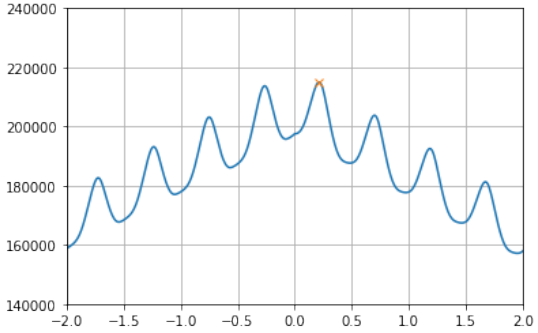
Once the time lag has been removed, a Pearson correlation coefficient was used to determine the strength of the linear relationship between the PPG and ABP signals which



(a) Plot of PPG and ABP to time



(b) Aligned PPG and ABP waveforms after removing PPG lag



(c) Cross correlation between PPG and ABP, with x at peak to indicate lag

Figure 2.7: Comparison between PPG and ABP before and after time alignment

was a good indication of the quality of both signals. This may seem redundant after performing cross correlation, but the Pearson correlation coefficient is advantageous in this case due to not taking amplitude scale into account [63]. It also has the benefit of verifying that the lag removal step happened correctly. The remaining 2434 patients with approximately 59,000 data entries in total had correlation coefficients from 0.65 to 0.99 with an average of 0.9. Eliminating those with a coefficient below 0.9 to ensure that only the best quality training samples were used left over 28,000 entries from 1673 patients.

BP selection criteria

The final step in eliminating abnormal data sets was verifying that the blood pressure fell within a realistic range. The criteria used for this are summarised in Table 2.2 and are the same as those used by Xing *et al.* [27]. The first criteria are the minimum and maximum allowed values for SBP and DBP. The next criterion is the maximum difference in DBP or SBP allowed between consecutive beats, which has to be less than double the standard deviation of the entire measurement. The difference between SBP and DBP must be more than 20 mmHg. Finally the period of the wave must be between 0.3 and 3 seconds, which equates to a heart rate of between 20 and 200 BPM.

The SBP and DBP for a sample were chosen as an average of the values at the systolic

Table 2.2: Selection criteria used to eliminate abnormal BP samples [27]

Feature	SBP	DBP	ΔSBP	ΔDBP	SBP-DBP	T
Abnormality criteria	$> 180mmHg$ or $< 80mmHg$	$< 20mmHg$	$> 2 \times \sigma_{SBP}$	$> 2 \times \sigma_{DBP}$	$< 20mmHg$	$< 0.3s$ or $> 3s$

and diastolic points across the 8 second sample. The systolic pressure is the highest point from a single period of the waveform where the diastolic pressure is the lowest point.

After all the selection criteria had been applied, there were approximately 8000 samples from 930 patients that could be used for training and testing.

Resampling, filtering and normalisation

A smartphone application would take measurements at 30 Hz, so the MIMIC PPG samples had to be resampled down from 125 Hz to 30 Hz. Since 125 is not a multiple of 30 the sampling points would not line up exactly. A linear interpolation method was used to calculate the values of the sampling points that did not align with the original signal's.

To further ensure that the signal measured from a smartphone and the test samples match as closely as possible, a first order Butterworth band pass filter was applied to both signal types. The low-pass filter was used to remove any sampling noise that may occur from the smartphone measurement, and has a cut-off frequency of 14 Hz, just below the Nyquist frequency which is 15 Hz. The high-pass filter was used to remove any low frequency slope or oscillations from slight movement of the user's finger, for which the cut-off frequency was chosen as 0.3 Hz to accommodate for slow pulses as well as the lower frequency components of the signal. This resulted in the following transfer function:

$$H(s) = \frac{\omega_{CH} s}{(s + \omega_{CL})(s + \omega_{CH})} \tag{2.33}$$

$$H(s) = \frac{28\pi s}{(s + \pi)(s + 28\pi)} \tag{2.34}$$

$$H(s) = \frac{87.96s}{s^2 + 91.1s + 276.35} \tag{2.35}$$

which, using the bilinear transform, led to the following output formula:

$$y[n] = 0.87952066x[n] - 0.87952066x[n-2] + 0.13003209y[n-1] + 0.75904131y[n-2] \tag{2.36}$$

A comparison between the original signal and the signal once it had been resampled and filtered as described above can be seen in Figure 2.8. This figure illustrates visibly that the main shape and structure of the waveform has remained intact after the resampling and filtering processes.

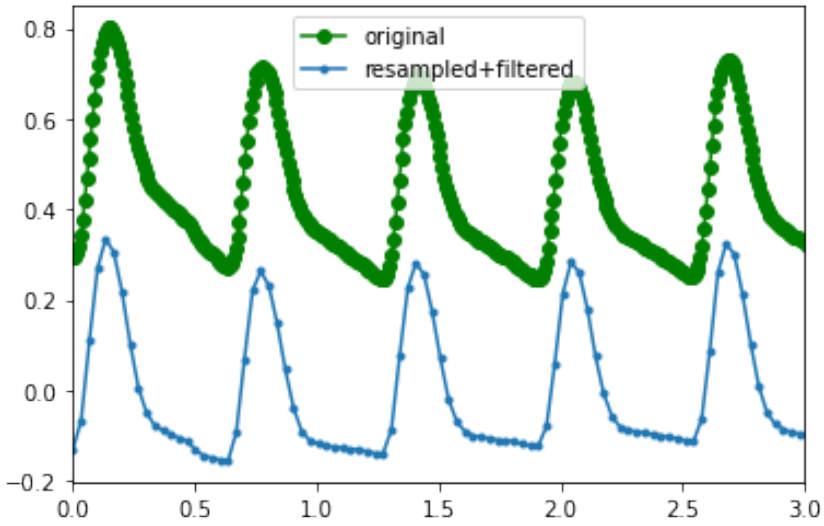


Figure 2.8: PPG waveform resampled to 30Hz and bandpass filter applied

Single frame isolation

A single frame from each sample was isolated to be used for feature extraction. One frame was defined to contain a complete cardiac cycle with 10% of the previous cycle and 5% of the following cycle, which was then padded with zeros to a length of 45 data points per frame which is equivalent to 1.5 seconds. The frame was then normalised to produce a sample as seen in Figure 2.9.

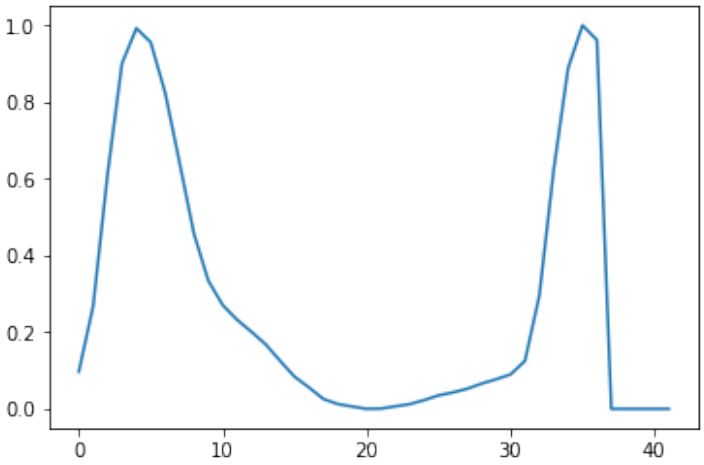


Figure 2.9: Single frame of resampled, filtered and normalised PPG waveform with zero padding to 45 samples

2.4.3 Feature extraction

As seen in Section 2.4.1, the relationship between blood pressure and PPG may be seen in the frequency response of the PPG. This implies that a FFT of the PPG wave may provide enough information to determine blood pressure. This is the route that Xing [27] followed with success. However, further consideration had to be made because of our much lower sampling rate. Figure 2.10 shows what an FFT of the PPG test data looks like.

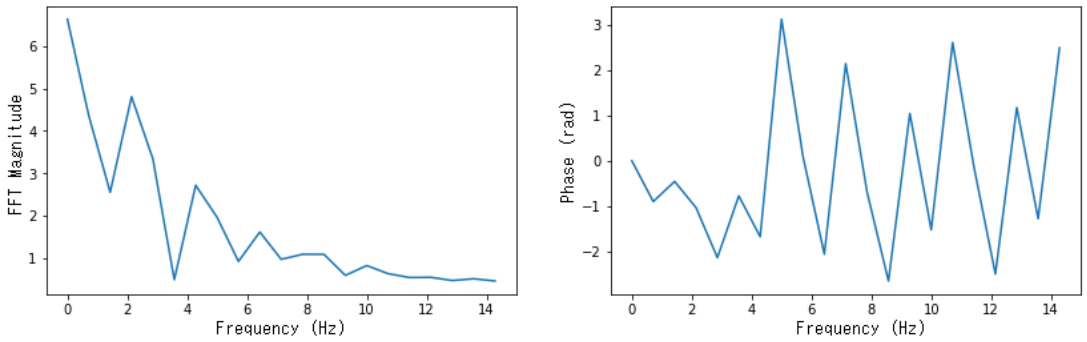


Figure 2.10: FFT magnitude (left) and phase (right) of single wave PPG

Considering that blood pressure is also a function of cardiac output and systemic vascular resistance, there are many time based features that would also be of value, such as systolic time (rise time which is dependent on RC), diastolic time (fall time which is dependent on RC), systolic peak, diastolic notch, systolic area, diastolic area, etc. Although all of these can be worked back from an FFT, some may require more complexity than others. A time based solution may fit more easily, possibly requiring a less complex network. Mousavi [29] had success with an approach of using all the points from a single wave of a PPG signal to train several different regression models.

For this solution, we tested a combined approach that used both frequency and time based inputs. The time based signal is the 45 data points frame seen in Figure 2.9. The frequency response was the first 14 points (10 Hz) of the magnitude and phase response seen in Figure 2.10. The reason for only using lower frequency points is because PPG and ABP correlate best at lower frequencies and start to diverge close to the cut-off frequencies of the low pass filter effects from arteriol and capillary resistance and elasticity.

2.4.4 ANN design

Test-train split

The MIMIC dataset was split up into three parts to be used for training, validation and initial testing. With 70% of the dataset (approximately 6650 samples) used for training and 15% (approximately 1400 samples) used for validation and testing each.

The distribution of the training data can be seen in Figure 2.11 with normal probability

plots for the same data in Figure 2.12. This was checked to ensure that it was close to a normal distribution to limit bias in the model. As can be seen from Figure 2.11, it seems that the diastolic pressure is biased slightly towards lower pressures. When looking at the SBP probability plot, the values close to the cut off points seem to approach a uniform distribution, but for the most part the data is close to a normal distribution. The probability plot of DBP approaches an inverted C-shape, indicating a right-skewed distribution as was noted from the histogram. This remained true for any subset of the data chosen and may be a result of the measurements being from ICU patients who were constantly sedentary.

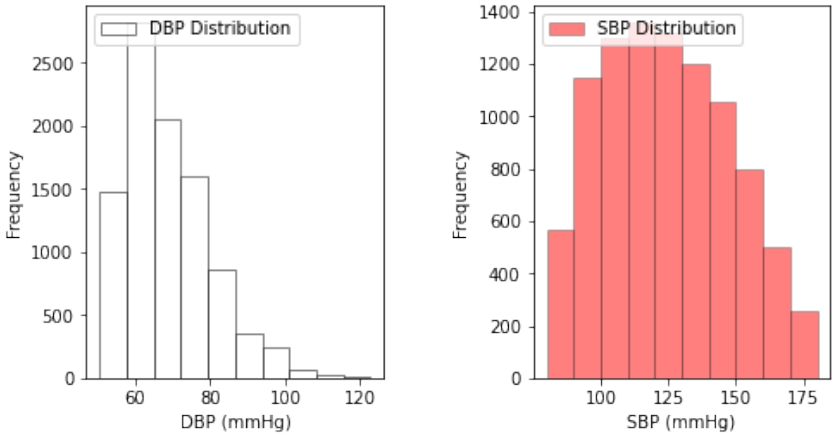


Figure 2.11: Distribution of training data diastolic pressure (left) and systolic pressure (right)

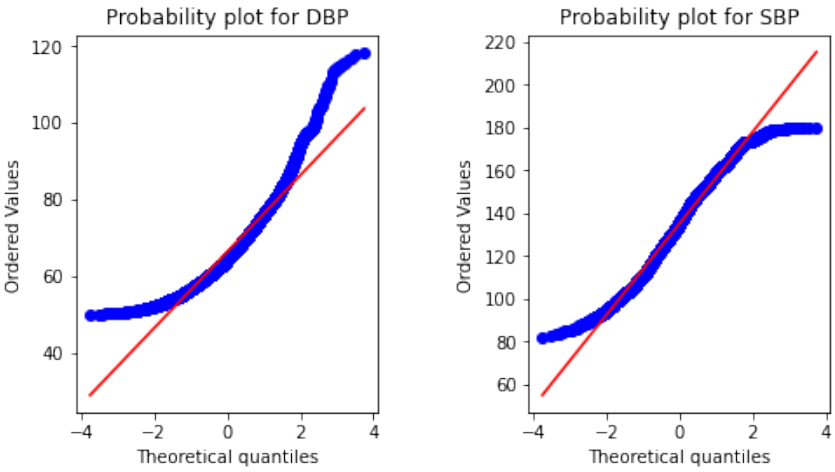


Figure 2.12: Normal probability plots of training data diastolic pressure (left) and systolic pressure (right)

Activation functions

The choice of activation function is very important. In the hidden layer it will control how well the network model learns the training dataset. The choice of activation function

in the output layer will define the type of predictions the model can make.

ReLU (Rectified Linear Unit) is common because it is both simple to implement and effective at overcoming the limitations of other previously popular activation functions, such as Sigmoid and Tanh. Specifically, it is less susceptible to vanishing gradients that prevent deep models from being trained, although it can suffer from other problems like saturated or "dead" units (specifically when dealing with negative values). Since, in this model, all input and output values will be larger than 0 the ReLU activation function would not be as susceptible to saturation and was a suitable choice for all hidden layers in this application.

For the output layer, because this is a regression problem, an ordinary linear activation function was used.

Number of layers/nodes

The exact number of hidden layers were largely determined by trial and error. Initially the number of nodes per hidden layer was chosen to be 60. Then this was tested with a different number of hidden layers and nodes. Finally, the best results were obtained at 4 layers of 120 nodes and resulted in a final network structure as seen in Table 2.3. The output layer contains two nodes, one for SBP and one for DBP.

Table 2.3: ANN design for blood pressure estimation

	Type	**Number of nodes**	**Activation function**
Input Layer	Dense	73	ReLU
Hidden Layer 1	Dense	120	ReLU
Hidden Layer 2	Dense	120	ReLU
Hidden Layer 3	Dense	120	ReLU
Hidden Layer 4	Dense	120	ReLU
Output Layer	Dense	2	Linear

Loss function

Since the network must provide a continuous value, and is therefore fundamentally a regression problem the main loss functions considered were MAE[7] and MSE[8].

MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average of the absolute differences between prediction and actual values. When applied to loss in regression problems however, it fails to punish large errors in prediction. MSE measures the squared average distance between the real data and the predicted data. Here, larger errors are well noted. However, it squares the units of data as well so may not be suitable when evaluating the performance of the model.

[7]Mean Absolute Error
[8]Mean Squared Error

For its stronger ability to pick up large errors, MSE was chosen as the loss function for this model. However, MAE was used in the evaluation phase to make better sense of the scale of the errors and to compare it to the measurement standards.

Optimiser

The optimisation algorithm used was the Adam optimisation algorithm, which is an extension to stochastic gradient descent and has become one of the most popular optimisers for deep learning applications [64]. In this case it was chosen due to its computational efficiency and quick convergence.

After 20 epochs of training, with a batch size of 4, the model achieved an accuracy of 94.8% on the training set and 93.4% on validation.

2.5 Chapter summary

In summary, this chapter discussed the design decisions that were made when developing the algorithms to measure a PPG signal using a smartphone camera and to determine heart rate and blood pressure from the PPG waveform.

The PPG measurements were made by determining a threshold based on the minimum and maximum values from the red channel during a calibration phase. The next step was determining the number of red pixels over each frame in the measuring stage, then normalising the inverted waveform to produce a normalised PPG signal.

HR was determined by using a bandpass filter on the PPG signal to only consider rates between 30 and 180 BPM, then to find the peaks on the waveform using the first derivative and measure the peak-to-peak time. This was done over several beats in a full measurement. Outliers were then removed before calculating an average.

The BP model was trained using clinical PPG and BP data from the MIMIC III database, by resampling PPG signals down to 30 Hz to match the frame rate from most smartphone cameras, applying a bandpass filter to remove sampling noise and low frequency movements then isolating a single waveform. An FFT was determined of the waveform, which was then used together with the time domain wave to train an ANN to determine SBP and DBP.

The procedure followed when testing the HR algorithm as well as the BP model is discussed in Chapter 4, with the in depth results and analysis of the testing procedure in Chapter 5.

CHAPTER

3

METABOLIC SYNDROME CLASSIFICATION

This chapter discusses the design of the ANN to be used to screen for MetS. In order to answer the questions posed in the problem statement, two separate models are built. The first model uses primarily quantitative inputs from features that are ordinarily used to determine MetS, excluding those features that cannot be measured using a smartphone or wouldn't be ordinarily known by the user. The second model includes lifestyle factors, such as activity level and smoking or drinking habits, with the features used in the first model. After looking at some of the relevant literature on the topic, a brief overview is given of the design considerations that need to be made with regards to the usability of the application's interface. With that in mind, the features are then selected, which are used to train and optimise the ANNs.

3.1 Relevant literature

There have been several studies using machine learning to identify subjects with MetS. One used an ANN, SVM[1], PCA, decision tree, random forest and association analysis for modelling and construction of predictive models for MetS characterisation, with varying accuracies, some being above 90% [14]. The research in [14] added a lot of value for determining the relationship between the different components of MetS.

[1]Support Vector Machine

Another study with a lot of relevance to this dissertation was conducted by Ivanović, where an ANN was used to diagnose MetS using only non-invasive, low-cost and easy to obtain parameters; namely gender, age, body mass index, waist-to-height ratio, systolic and diastolic blood pressures [15]. This research is promising as it has been shown that a reliable classifier can be designed using easily obtainable information that can easily be entered on a smartphone, with the exception of blood pressure.

In a similar study, ANN and multiple logistic regression was employed for identifying MetS in patients treated with second-generation anti-psychotics [65]. The results indicated that ANN and logistic regression models gave high accuracy of 88.3 and 83.6%, respectively, with waist circumference, BMI, DBP and gender being important variables for identifying MetS.

The study by Ushida [12] found that alcohol and cigarette use highly correlated with the most significant risk factors for the development of metabolic syndrome. Another study that may have showed the most promise with regards to information attainable form lifestyle factors is one by Jahandideh, that used an ANN to determine cardiovascular disease risk factors (HDL cholesterol, total cholesterol, triglycerides) from factors such as sex, age, build, weight, marital status, individual's status in the family, physical activity, hours of sleep per day, smoking, tobacco type and BMI [47]. The ANN delivered high accuracy for detecting triglycerides (85.75%), HDL cholesterol (91.25%) and total cholesterol (93.75%) [47]. The study also used a linear regression model, that did not have the high accuracy of the ANN but was more useful than the ANN in determining which variables correlate the most [47]. The study found a strong influence of the sex variable on cholesterol, HDL cholesterol and triglycerides; of BMI on triglycerides; and weight and tobacco kind on HDL cholesterol [47]. This study shows, therefore, that it may be possible to evaluate these non-invasive measurements or their effects from lifestyle factors, implying that these lifestyle factors could be used as predictors for MetS as well.

From the above-mentioned literature, it was evident that an ANN could be suitable to detect MetS from a reduced feature set. The research by Ivanović [15], which used an ANN to determine MetS from only non-invasive features, is used as a starting point. There are some departures from the existing research based on the features available in the training data, and the addition of other lifestyle-based features are also tested to determine whether it would make a significant impact on the accuracy of the model.

3.2 Application interface

As mentioned in chapter 1, the ideal medical screening test should be accessible, simple to use, inexpensive, and associated with minimal discomfort to the person/s being screened [31]. This means that, in a mobile application, an easy to use and intuitive user interface is important for it to be considered an effective screening solution. A certain amount of feedback may also be required from the application to ensure that a user is on the right track.

As determined in the previous chapters, a small portion of the information required to

determine if an individual has MetS would already be known by the individual (such as weight, height, etc.). The remaining information required would either need to be measured or estimated from a PPG signal. Therefore, the inputs available was limited to user input fields and standard smartphone sensors, more specifically a cellphone camera. Based on the promise shown in previous studies that determined MetS using only non-invasive measurements [15], a similar technique was employed in this dissertation.

User interaction and usability were major deciding factors for which features to keep and which to discard. If it were determined that a feature may be of value but would greatly reduce the usability, and there was no way of reducing the feature's complexity, it was discarded. As, for example, was done with calorie consumption or fat percentage discussed in the introduction. TEE[2], as will be seen in Section 3.4.3, is an example of a feature that was included after having complexity reduced to an extent that it would not discourage users from using the application.

The user's interaction with the application can be broken down into three areas, namely:

1. Data entry
2. Measurement
3. Results.

The data entry can be split into biometric data (length, weight, waist circumference, etc.) and biographical data (medical history, eating, drinking and smoking habits, etc.). Biographical data is simply referred to as "lifestyle factors" going forward. Measurement will include all measurements to be taken by the smartphone's sensors, in this case PPG measurement. Results is the section where the user can view all their measured results and final MetS result, which could be displayed as either a binary prediction or a likelihood based on the model's confidence.

Figure 3.1 shows an example of what the proposed interface may look like, based on a POC[3] Android application that was developed for testing purposes.

3.3 Data analysis

The dataset used to create the metabolic syndrome classifier is the SAPBA dataset collected by researchers from the North West University [46]. It contains over 1200 fields and measurements for 409 participants. Many of these fields are the same conditions measured with different techniques or different equipment. Since these different measurement techniques fall outside the scope of this study, only a single relevant measurement for each condition was included. Measurements that could not be feasibly collected on a smartphone and do not apply to any of the standard definitions of MetS (discussed in

[2]Total Energy Expenditure
[3]Proof of Concept

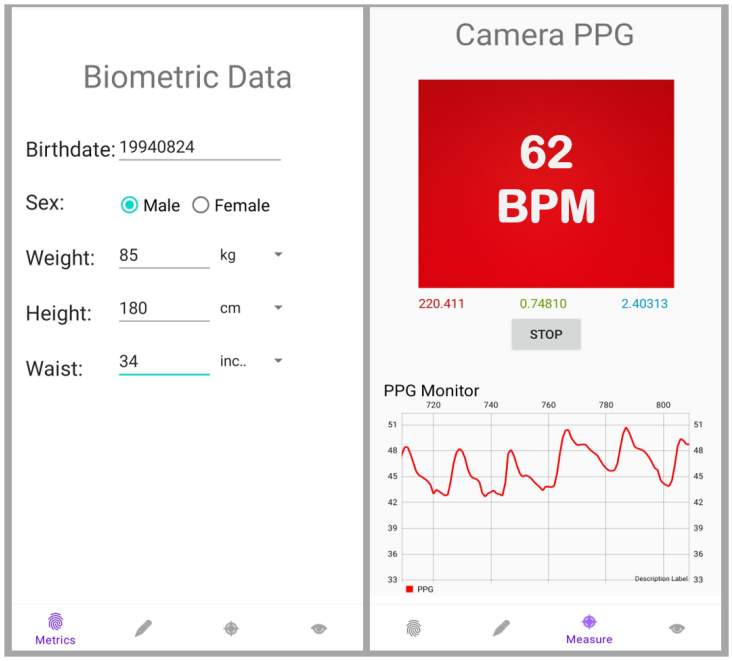


Figure 3.1: Example of biometric data and PPG measurement sections from the application, with bottom navigation bar

Section 4.2.2) were also excluded. If any of the applicable fields for a participant was blank the participant was excluded. This left 24 applicable fields for 402 participants. A description of every field and the method used to measure it is provided in appendix A.

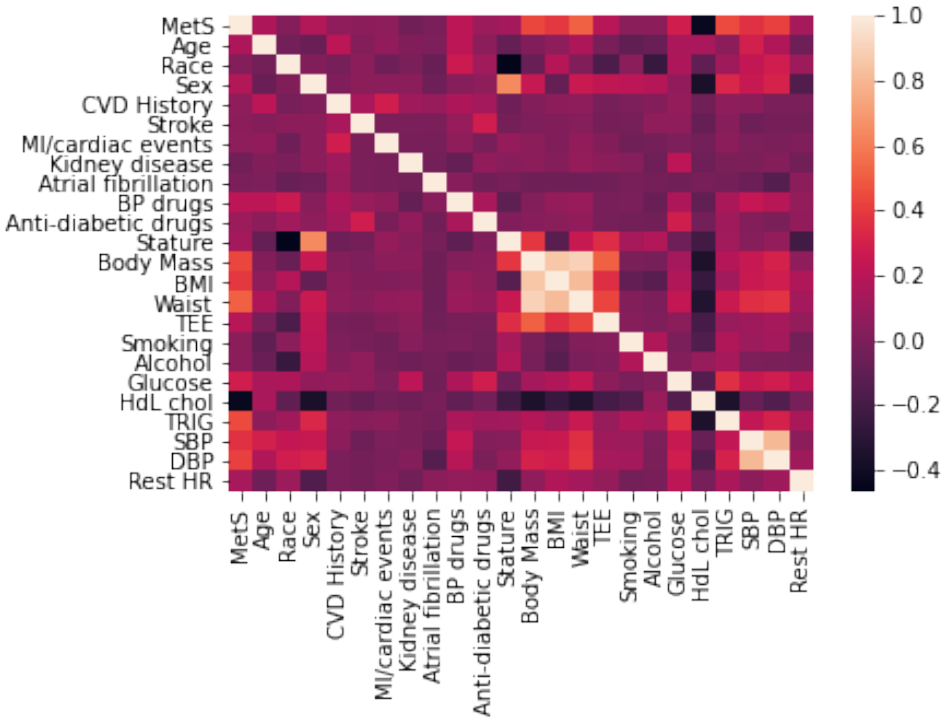


Figure 3.2: Correlation matrix for data from SAPBA dataset

The correlation matrix of all the applicable variables in the SAPBA dataset can be seen in Figure 3.2. The correlation between variables gives a good initial indication of correlation that may be between variables. Lack of correlation does not necessarily rule out a relationship though. It may just indicate that, if there were a relationship, it may be non-linear. It is because of this non-linear relationship with many of the available inputs that an ANN will be used to determine MetS, since neural networks are particularly well suited to non-linear classifications.

As can be seen from Figure 3.2, the variables that correlate the most with MetS are (in order of highest absolute correlation to lowest):

1. Waist circumference (0.51)
2. Body mass (0.47)
3. Serum triglyceride levels (0.45)
4. Serum HdL cholesterol (0.43)
5. Diastolic blood pressure (0.42)
6. BMI (0.39)
7. Systolic blood pressure (0.36)
8. Blood glucose (0.30)

As would be expected, the values that correlate the closest to MetS are those that are used in ordinary circumstances to make a diagnosis. All of these have a correlation between 0.3 and 0.6, which is a significant correlation but not enough to make a valid diagnosis on its own. The high correlation of body mass and waist circumference is promising, since both of these values are provided by the user on the application.

In Figure 3.3 the distribution of some of the features for patients with and without MetS can be seen. Comparing the two distributions for each feature may provide some insight into how the feature relates to MetS. From Figure 3.3, it can be seen that below the age of 40-45 the likelihood of having MetS is lower, but over the age of 45 the distribution is very similar. BMI, HDL cholesterol and SBP all have close to normal distributions with just a shift in mean between positive and negative MetS diagnoses. Specifically, there is a mean difference in BMI of about 5 with positive diagnosis being higher, negative diagnosis shows an average HDL that is 0.6 mmol/l higher than positive diagnosis, and a difference in SBP of about 10 mmHg. These differences are significant, but not to such an extent that a single feature could be used to accurately screen for MetS.

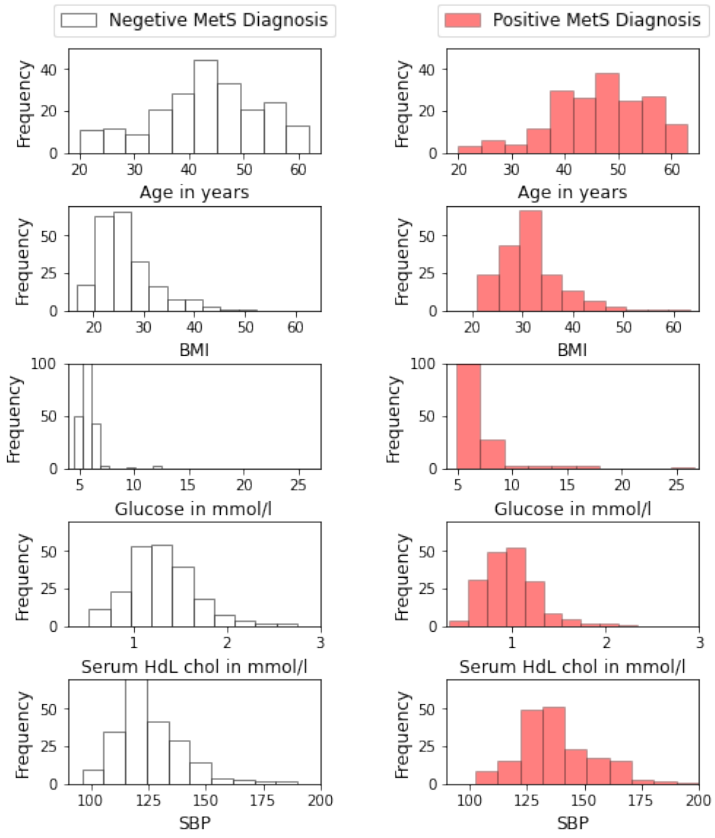


Figure 3.3: Distribution of metabolic syndrome risk factors

3.4 Preprocessing and feature selection

3.4.1 Main features

The initial features chosen to determine MetS was largely based on [15] and the features that were available in the SAPBA dataset. However, since the training data used in this dissertation is different to that used by [15], the results may also differ somewhat. The following features were chosen for the first model:

1. Age
2. Gender
3. BMI
4. Waist-to-height ratio
5. SBP
6. DBP
7. Heart rate

These features were chosen because they are all non-invasive measurements that can potentially be made by a smartphone or because they would already be known by a user.

3.4.2 Lifestyle factors

A second model was built that included all the other information that was available in the SABPA dataset that may be useful for determining MetS. The following features were selected and included with the previous features:

1. Medical history
2. Alcohol use
3. Smoking
4. Activity level

Both medical history and activity level aren't directly available in the dataset and required some preprocessing to determine. Activity level is discussed in Section 3.4.3. Medical history is a combination of the diseases provided in the SAPBA dataset, such as diabetes or stroke, by means of an OR operation. The conditions that were included in the OR operation are:

- Cardiovascular disease history
- Stroke history
- Myocardial infarction/ cardiac events history
- Kidney disease history
- Atrial fibrillation
- Use of anti-hypertensive drugs
- Use of anti-diabetic drugs

Note that while some of these features (like medical history) may not necessarily be lifestyle related, for the sake of simplicity going forward, when differentiating between the two models the inclusion of lifestyle factors is the distinction that will be made.

3.4.3 Activity level

Several studies have shown the impact of exercise on MetS [7][3][17]. The SAPBA dataset does not include any direct indicator of activity level, however, it does include TEE (Total Energy Expenditure). This is a measure of the total caloric consumption of an individual in a day, which was measured using an activity tracker. The BMR[4], which is the amount of calories required per day without any physical activity, can be predictively calculated with the revised Harris-Benedict equations [66]:

Table 3.1: Harris-Benedict equations as revised by Mifflin and St Jeor

Men	BMR = (10× weight in kg) + (6.25× height in cm) + (5× age in years) + 5
Women	BMR = (10× weight in kg) + (6.25× height in cm) + (5× age in years) - 161

The activity level is then determined by looking at the ratio of TEE over BMR. The following classification is defined by the original Harris-Benedict equations [66]:

- Little/no exercise: $TEE/BMR = 1.2$
- Light exercise: $TEE/BMR = 1.375$
- Moderate exercise (3-5 days/wk): $TEE/BMR = 1.55$
- Very active (6-7 days/wk): $TEE/BMR = 1.725$
- Extra active (very active & physical job): $TEE/BMR = 1.9$

With this classification it will be easy to link an activity level to users using the application. In the model the TEE / BMR ratio will be normalised, so that each option a user could fill in would represent the following in the model:

- Little/no exercise: 0
- Light exercise: 0.25
- Moderate exercise (3-5 days/wk): 0.5
- Very active (6-7 days/wk): 0.75
- Extra active (very active & physical job): 1

3.4.4 Scaling

All of the features were scaled to be between 0 and 1. Some were binary values, which were either 0 or 1. Activity level has five possibilities ranging from 0 to 1. Analog values were scaled to be between 0 and 1 using min-max scaling per feature of the training set. A description of each feature's value range is given in Table 3.2.

[4]Basal Metabolic Rate

Table 3.2: Value ranges of metabolic syndrome features

Feature	Value range
Age	0.0-1.0
Gender	0 or 1
BMI	0.0-1.0
Waist-to-height ratio	0.0-1.0
SBP	0.0-1.0
DBP	0.0-1.0
Heart rate	0.0-1.0
Medical history	0 or 1
Alcohol use	0 or 1
Smoking	0 or 1
Activity level	0.0-1.0

3.5 ANN design

3.5.1 Network Optimisation

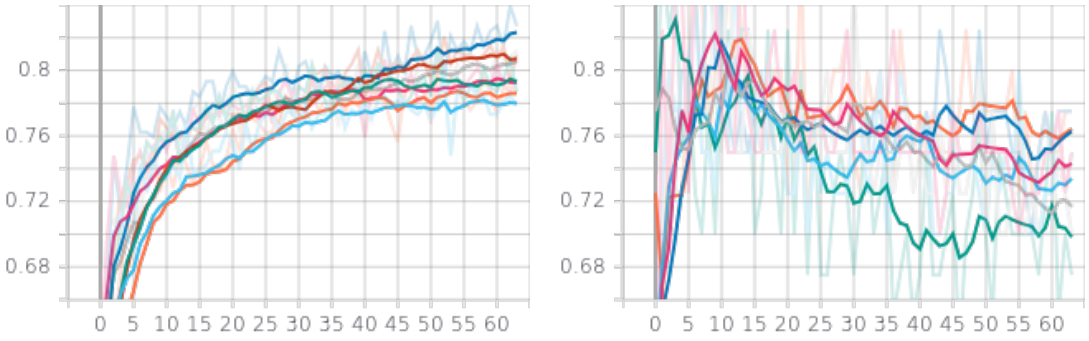
As with the BP model, in this model, all input and output values are between 0 an 1, so the ReLU activation function was used for all hidden layers. For the output layer, because this is a single variable classification problem, a sigmoid activation function on a single neuron layer was the ideal solution.

In order to get an observable indicator of how well optimised the ANN topology was optimised to the features, TensorBoard was used to visualise the performance of multiple ANN topologies over multiple epochs. This was especially important in this case, because the low number of training samples meant overtraining was a problem likely to be encountered. The number of nodes per hidden layer was tested in factors of 10, i.e. 10, 20, 30. And this was tested for both one and two hidden layers.

Figure 3.4 shows the results from one of the sessions that was run on the MetS model exluding lifestyle factors. This session ran for 64 epochs, which was chosen as an arbitrary high starting point to get a large scale view of what the performance could look like. Similar sessions were run multiple times to reach a conclusion on what the best topology would be for our case.

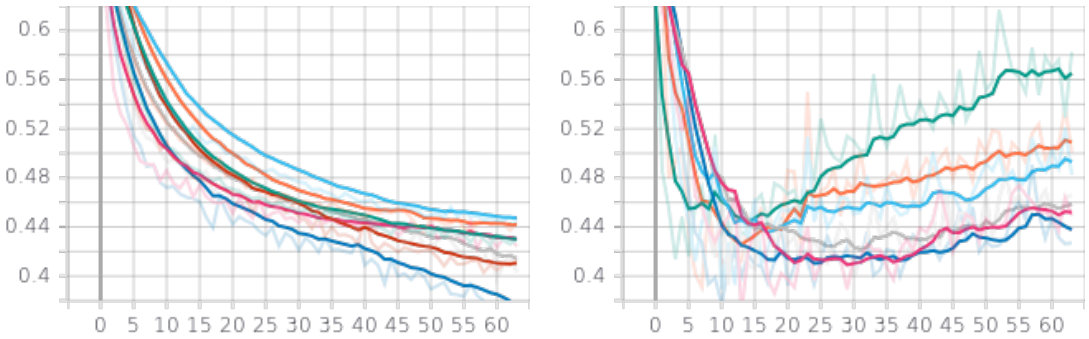
Although the details in these figures are harder to distinguish without the interactive features of TensorBoard, Figures 3.4 and 3.5 were included to give an indication of what the observations were in the optimisation period. One general observation that is immediately apparent from Figures 3.4b and 3.4d was that the drop in validation accuracy and rise in loss meant that overtraining became an issue after 15 to 20 epochs. This tended to only happen later for the model including lifestyle factors, as is somewhat apparent from Figures 3.5a and 3.5b.

The green, orange and light blue curves in the loss function represented the models that



(a) Training set accuracy

(b) Validation set accuracy

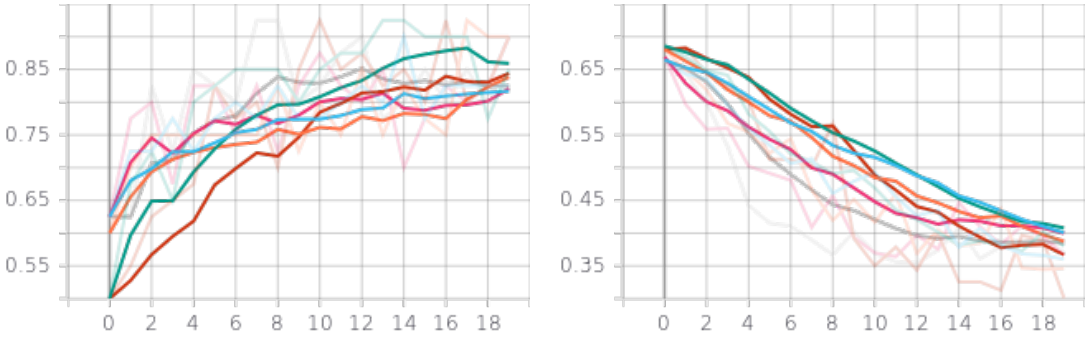


(c) Training set loss

(d) Validation set loss

Figure 3.4: Performance of basic MetS models on validation and test sets over 64 epochs for different ANN topologies

had 2 layers, indicating that a deep network would not be ideal for this dataset. While the results from other sessions were not as clear cut in this regard, the trend tended to stay the same. The trend with regards to layer count was not always as clear in the case of the model including lifestyle factors, as can be seen in Figure 3.5 but the single layer network tended to have slightly more consistent performance. Using similar topologies for the two models, also helps to ensure that performance differences seen are due the added features and not network performance.



(a) Validation set accuracy

(b) Validation set loss

Figure 3.5: MetS models with lifestyle features on validation sets over 20 epochs for different ANN topologies

Finally, the number of neurons for the hidden layer was chosen as 30 in both cases. While

the performance remained very similar regardless of the number of neurons chosen, this option gave the most consistent results over multiple tests on both models.

3.5.2 Without lifestyle factors

A summary of the ANN topology that was optimised for the MetS model without lifestyle factors can be seen in Table 3.3.

Table 3.3: ANN topology for Mets **without** lifestyle factors

	Type	**Number of nodes**	**Activation function**
Input Layer	Dense	7	ReLU
Hidden Layer	Dense	30	ReLU
Output Layer	Dense	1	Sigmoid

3.5.3 With lifestyle factors

A summary of the ANN topology for the MetS model with lifestyle factors can be seen in Table 3.4. The model was chosen to match the model excluding lifestyle factors as closely as possible, without sacrificing accuracy.

Table 3.4: ANN topology for Mets **with** lifestyle factors

	Type	**Number of nodes**	**Activation function**
Input Layer	Dense	11	ReLU
Hidden Layer	Dense	30	ReLU
Output Layer	Dense	1	Sigmoid

3.6 Chapter summary

In this chapter, the design decisions that were made when developing the algorithms to screen for MetS were discussed. The focus was on achieving this with limited features that could be available to a smartphone user or could be measured using a smartphone.

User interaction and usability were major deciding factors for which features to keep and which to discard. A portion of the information required to determine MetS that would already be known by the individual were chosen, such as weight, height, etc. The remaining features, such as HR and BP, would be determined from the PPG measurements in the previous chapter.

The data used for training and optimisation was from the SAPBA dataset [46]. From this, the following features were selected:

1. Age
2. Gender
3. BMI
4. Waist-to-height ratio
5. SBP
6. DBP
7. Heart rate
8. Medical history
9. Alcohol use
10. Smoking
11. Activity level

Two models were developed, one excluding and one including lifestyle factors (features 8 to 11 in the list above). All the input features were scaled to be between 0 and 1, and an optimised ANN topology was determined for each model. The procedure followed when testing these models is discussed in Chapter 4, with the results from the testing procedure in Chapter 5.

CHAPTER

4

EXPERIMENTAL PROCEDURE

In this chapter the experimental procedure used to validate the solutions designed in chapter 2 and 3 is discussed. The solutions are assessed according to how well they answer the research questions posed by the problem statement. The methods used to obtain data and how the data is used are also discussed.

4.1 Addressing the research questions

The following questions were formed in Chapter 1 to expand on the main research question:

1. Can a smartphone be used as a plethysmograph in order to accurately determine heart rate and blood pressure?
2. Is it necessary to determine all the characteristic risk factors for MetS (abdominal obesity, fasting blood sugars, serum triglyceride levels and blood pressure) in order to diagnose metabolic syndrome?
3. Can information from lifestyle factors, such as diet, exercise or smoking, be used to improve the accuracy of MetS predictions?
4. Would an application of this kind be easy enough to use for it to be a feasible solution?

These questions are used as a guide to answer the research question, and by answering them, we hope to adequately address the overall problem.

4.1.1 Answering question 1: PPG based heart rate and blood pressure

There have been previous studies that determined blood pressure from PPG data obtained from blood oximeters. There are also many heart rate monitoring cell phone applications already available on the Google Play Store as well as the Apple App Store, some of which have also been the subjects of research studies [21], [26], [37]. However, there have been fewer studies to show similar BP results with PPG data obtained from a cell phone camera. There are also no known online datasets available that contain blood pressure and PPG data, measured using a cell phone camera. For this reason it was necessary to obtain data measured with a smartphone that could be used to test the proposed heart rate and blood pressure algorithms.

For the initial training, testing and validation the MIMIC III waveform database was used [62]. This model was then integrated into an application that would be used to gather cellphone based BP, HR and PPG data from participants. The data collection process is discussed in Section 4.3

4.1.2 Answering questions 2 and 3: Metabolic Syndrome

As discussed in the previous chapter, to determine the impact of removing some elements from the consideration when determining MetS, the following features were chosen from the SABPA dataset to use in an initial model:

1. Age
2. Sex
3. BMI
4. Waist-to-height ratio
5. SBP
6. DBP
7. Heart rate

These features were chosen because they are all non-invasive measurements that can potentially be made by a smartphone or would already be known by a user.

A second model was built that included all the other information that was available in the SABPA dataset that may be useful for determining MetS. The following features were selected:

1. Medical history

2. Alcohol use
3. Smoking
4. Activity level

Both of these models were evaluated in the same manner, discussed in Chapter 5, and compared to other standard definitions of MetS. To further test the potential accuracy of a fully integrated application, the distribution of the errors from the BP and HR results were used as margins to add random errors to the testing sets of the MetS model, including lifestyle factors.

4.1.3 Answering question 4: Usability

Observable and quantifiable metrics were determined based on the ISO 9241-11 standard, which defines usability as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* [48]. The usability of the application was measured in terms of effectiveness, efficiency and satisfaction.

There was, however, a major limiting factor with regards to determining the usability of the application. Since the data collected in the data gathering process only included PPG, HR and BP measurements, participants did not get a perspective of the entire application, but rather just a single aspect thereof, i.e. the PPG measurement. While this means that the results will only be limited to a single element of the application, considering that it is the most time consuming, involved and experimental aspect of the application (the rest simply entails filling in text fields), this should still provide a valuable indication of what the usability of a full MetS application would be.

4.2 Overview of the datasets for testing

There were three different datasets used for testing the different algorithms and classifiers, which are summed up in Table 4.1. In this section we will look at those datasets in more depth and discuss how each was used to help answer the research questions.

4.2.1 MIMIC

The dataset that was used for training, validation and initial testing of the blood pressure algorithm is the MIMIC III Waveform Database [39], [62]. The original database contains 67,830 record sets for approximately 30,000 ICU patients, usually including signals such as ECG, ABP, PPG, respiration and more.

The data sifting and preprocessing are all discussed in detail in Chapter 2. After all the selection criteria had been applied, there were approximately 8000 PPG and BP samples

Table 4.1: Summary of the three datasets used for testing

Dataset	**# Records**	**Uses**
MIMIC III	8000 (1200 for testing)	BP model training and testing
SABPA	402	MetS models training and testing
Smartphone	133	- HR algorithm testing - BP model testing - BP and HR errors for MetS model testing - Usability testing

from 930 patients that could be used for training and testing. The remaining data was split up into a ratio of 70/15/15 for training, validation and testing. This resulted in a test dataset of approximately 1200 records, which is processed for results in Chapter 5.

The same standards used to validate the accuracy of standard cuff-based shygmomanometers was used for the blood pressure algorithm on the test set. Standardised protocols for validating the clinical accuracy of non-invasive shygmomanometers have been available since 1987, some of which were developed by standards bodies and others by professional organisations [67]. A 2020 report by the World Health Organization [67] listed these standards as follows:

- Association for the Advancement of Medical Instrumentation (USA), 1987, 1992 and 2002
- British Hypertension Society, 1990, 1993
- German Hypertension League (Deutsche Hochdruckliga), 1999
- European Society of Hypertension, 2002, 2010
- ISO, 2009
- American National Standards Institute, Association for the Advancement of Medical Instrumentation and International Organization for Standardization, 2009, 2013
- Association for the Advancement of Medical Instrumentation, European Society of Hypertension and International Organization for Standardization, 2018

The protocols differ significantly with one key difference being the number of participants required for testing. A joint report from 2019 by the ESH[1], the AAMI[2] and ISO[3] endorsed the ISO 81060-2; 2018 standard and called it the single universal standard that will replace all other previous protocols [68]. For this reason, the ISO 81060-2 standard was used for validating the blood pressure algorithm.

[1]European Society of Hypertension
[2]Association for the Advancement of Medical Instrumentation
[3]International Organization for Standardization

4.2.2 SABPA

The SABPA dataset was collected by researchers from the North West University and originally used to identify factors leading to early vascular abnormalities in Sub-Saharan Africans [46]. The dataset contains over 1200 fields and measurements for 409 participants. Many of these fields are the same conditions measured with different techniques or different equipment. Since these different measurement techniques fall outside the scope of this dissertation, only a single relevant measurement for each condition was included. Measurements that could not be feasibly collected on a smartphone and do not form part of the definition of MetS were also excluded. If any of the applicable fields for a participant was blank the participant was excluded. This left 24 applicable fields for 402 participants.

The diagnostic method used by the SABPA researchers to determine MetS was similar to the NCEP ATPIII method, with slight modifications to some of the threshold values based on the South African standards where the research was conducted. Specifically, participants were marked positive for MetS when three of the following criteria were satisfied:

- Waist circumference $>$ 94 cm fom men or $>$ 80 cm for women (based on IDF standard for Sub-Saharan Africa)
- Blood glucose $>$ 5.6 mmol/l
- Sphygmomanometer BP: SBP $>$ 135 and/or DBP $>$ 85 mmHg
- HDL $<$ 1.036 mmol/l for men or $<$ 1.295 mmol/l for women
- Triglycerides $>$ 1.7 mmol/l

As discussed in Chapter 3, and mentioned in Table 4.1, the SABPA dataset was used to train the MetS models. It was also used for testing and data analysis purposes, as will be discussed in detail in Chapter 5. In order to have more baselines to compare the results from the MetS model to, two other methods of clinically diagnosing MetS were used, namely the IDF and WHO methods, as an indication of the variation that may be expected between diagnostic methods. The IDF method used exactly the same threshold values as used by SABPA, the only difference being that abdominal obesity (i.e. waist circumference) was a required parameter for a positive diagnosis. For the WHO method, insulin sensitivity was required, which was determined in this case by either high blood glucose or the use of anti-diabetic medication. Abdominal obesity for the WHO method was determined by either BMI ($>$30) or a waist-to-height ration of more than 0.9 for men or 0.85 for women. Table 4.2 shows the confusion matrices from these methods compared to the method used in the SABPA data.

Based on the values from Table 4.2, the IDF method had an accuracy of 95.5%, with a sensitivity of 90.2% and a specificity of 100%. The WHO method achieved an accuracy of 87.4%, with a sensitivity of 78.2% and a specificity of 95% when compared to the

Table 4.2: Confusion matrices of IDF and WHO MetS diagnoses compared to the method used in the SABPA dataset

		IDF		WHO	
		Positive	**Negative**	**Positive**	**Negative**
SABPA	**Positive**	166 (41%)	18 (4.5%)	144 (35.6%)	40 (9.9%)
	Negative	0	220 (54.5%)	11 (2.7%)	209 (51.7%)

SABPA method. These values indicate a standard variation that can be expected between diagnostic methods and will be used as a baseline in the coming chapters when discussing the effectiveness of the MetS algorithm.

4.2.3 Smartphone data

In order to verify the accuracy of the results obtained from the MIMIC dataset, real-world smartphone PPG data was required with corresponding BP measurements. This data had to be collected, because no such dataset is available at the time of writing. A study procedure was formed, which is discussed in Section 4.3. The data obtain was used for testing the HR algorithm, the BP model, generating BP and HR errors for testing the MetS model, and testing the usability of the application.

The study was approved by the NWU-Health Research Ethics Committee of the Faculty of Health Sciences of the North-West University before the data collection commenced and during the research all possible steps were taken to ensure that it would be conducted according to the ethical guidelines and principles of Ethics in Health Research: Principles, Processes and Structures [69].

4.3 Data collection process

This section discusses the process followed to collect real world data of smartphone PPG and BP measurements. The objectives of the data gathered were as follows:

- Test whether PPG data obtained from a smartphone camera can be used to determine heart rate and blood pressure.
- Determine if the HR and BP measurement was accurate enough to give a reliable metabolic syndrome result.
- Determine if the application is usable enough to be a feasible solution.

4.3.1 Data collection methodology

An Android application was created specifically for testing the PPG and blood pressure measurements. The application was developed to record the reference blood pressure and heart rate, PPG signal, participant feedback, calculate the heart rate and blood pressure from the PPG wave, and used a Google Apps Script API to backup all data to a Google Drive directory.

A commercial cuff-based sphygmomanometer (Omron HEM-712C [70]) was used to measure the participants' blood pressure. The measured value was entered into the application which saved it with the rest of the data on the cloud. A screenshot of this step can be seen in Figure 4.1a.

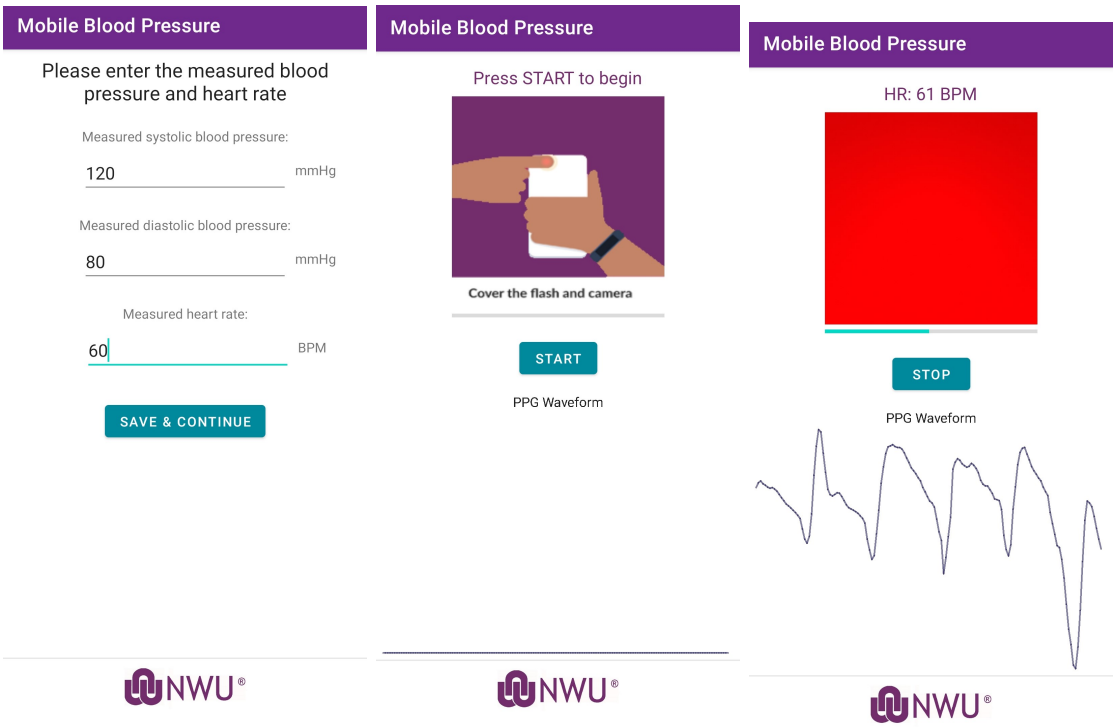
The participants were then given the smartphone with the application, on which they followed the instructions to take the PPG measurement. The same smartphone, a Samsung Galaxy A50, was used for all participants' measurements. Figures 4.1b and 4.1c show the screens that the participants would see while taking their measurement. During this process, besides the PPG measurement, the application would also measure the total time taken to perform a successful measurement and the number of attempts needed to get a successful measurement to be used in the usability analysis (efficiency and effectiveness). The timer started as soon as the activity opened and ended once a full signal could be measured, analysed and saved. The number of attempts were determined by counting every time the user pressed the start button, as this had to be done to restart the measurement every time the application went into an error state.

After a successful measurement the results would be shown, as seen in 4.1d and after pressing "Save & Continue", participants answered the questions used to rate the usability (user satisfaction).

Over the whole process the following data was collected from participants:

- Blood pressure measured with a cuff-based sphygmomanometer and typed into the application
- Heart rate measured with the sphygmomanometer and typed into the application
- Raw PPG signal data measured with the smartphone camera
- Heart rate determined by the application
- Blood pressure determined by the application
- Number of measurement attempts automatically counted by the application
- Time taken for successful measurement automatically measured by the application
- Application usability ratings given by participant, recorded on the application

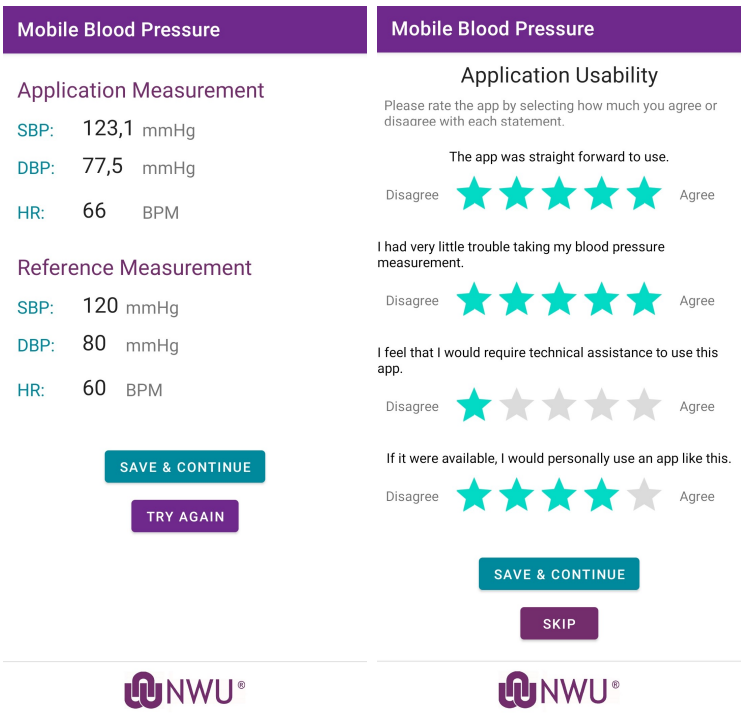
All of the information was recorded by the application and automatically uploaded to a Google Drive directory, from which further processing and statistical analysis could take place.



(a) Saving reference measurement

(b) PPG measurement start view

(c) PPG measurement



(d) Results view

(e) User rating page

Figure 4.1: BP data collection application activities

4.3.2 Study location

The study was conducted at Cachet Park in Potchefstroom, which was chosen due to its proximity to the gym facilities and surrounding restaurants, allowing for easy recruitment of participants. Some downsides to this location that were encountered during testing were a lack of control of conditions such as wind and light, and a lack of consistency with regards to participants' rest/ exercise status. The inclusion of post-exercise participants was initially seen as a benefit as it would provide a greater spectrum of data and provide more information when processed. However, because a sphygmomanometer was used for BP and HR measurement that meant that the two HR measurements could not happen simultaneously and enough time would pass between measurements for a drop to be seen in post-exercise HR. Nevertheless, while many of these factors would result in reduced test accuracy, they would provide a more realistic picture of what real-world use may look like.

4.3.3 Study population

The sample size was determined according to the ISO 81060-2 standard for non-invasive sphygmomanometers, which requires a sample size of at least 85, however, there were a total of 133 participants that took part in the study.

Due to the relationship between hypertension, sex and age, a study population was required that spanned a large age group with male and female participants. Possible changes in light absorption depending on skin tone meant that a variation in skin colour was also required. While none of these variables were recorded, during testing the we ensured to include as great a variation in participants as possible, spanning ages from 18 to over 70, including male and female participants, with multiple different skin colours. The proportions of these variables somewhat matched the demographic of the area, meaning, the majority of participants were ages 18 to 25, there were slightly more female participants than males, and approximately 80% of participants were white and the rest were black.

4.3.4 Study limitations

The focus of the data collection was to determine blood pressure using PPG from a smartphone camera, not to determine metabolic syndrome. Due to the focus on blood pressure measurement, other risk factors of metabolic syndrome were not measured. While the error found in the blood pressure and heart rate measurement were accounted for when assessing the model, this may still have an influence on the results for determining metabolic syndrome.

There are many aspects that may influence short term blood pressure, such as exercise, sleep and stress that was not controlled during this study. The key focus was just determining if the application BP and HR matched the reference.

4.4 Chapter summary

The goal of this chapter was to discuss the experimental procedure used to validate the solutions designed in chapter 2 and 3 based on how well they answer the research questions. The test data used for this purpose was also described.

The first question dealt with determining HR and BP using a PPG measurement. Two datasets would be required to answer this question. For the initial training, testing and validation the MIMIC III waveform database was used [62]. This model was then integrated into an application that would be used to gather cellphone based BP, HR and PPG data from participants.

The second and third questions dealt with the ability to screen for MetS without the use of the standard risk factors and including lifestyle factors. The MetS models, developed in the previous chapter, were evaluated using the SABPA dataset, and compared to other standard definitions of MetS. To further test the potential accuracy of a fully integrated application, the distribution of the errors from the smartphone BP and HR results is used in Chapter 5 as margins to add random errors to the testing sets of the MetS model.

The fourth question dealt with the usability of an application. The ISO 9241-11 standard was used, which defines usability in terms of effectiveness, efficiency and satisfaction. The data used for testing usability was collected with the PPG and BP measurements in the smartphone data collection process.

An Android application was created specifically for capturing the PPG and blood pressure measurements in the data collection process. The application was developed to record the reference blood pressure and heart rate, PPG signal, participant feedback and calculate the heart rate and blood pressure from the PPG wave. There were 133 participants that took part in this data collection process.

CHAPTER

5

DATA ANALYSIS AND RESULTS

This chapter discusses the results obtained from the experimental procedure described in the previous chapter. The results are divided according to heart rate, blood pressure, MetS and usability.

5.1 Heart rate

To assess the error of the heart rate algorithm, ME[1], MAE, and MAPE[2] were used. In line with prior research regarding PPG based heart rate devices [71]–[73] as well as recommendations from the AAMI, the Consumer Technology Association [74] and ANSI[3] [75], an acceptable error rate for the heart rate algorithm was defined to be 10%, as this is considered an accurate threshold for medical ECG monitors. This standard is more lenient than those used in some prior research, which used 5% [76], but is sufficient to use as baseline in this dissertation since the main goal is ensuring that the error is small enough to have a negligible influence on the MetS algorithm.

The participant data captured using the smartphone application during the data collection process was used to test the heart rate algorithm. As with the blood pressure algorithm, a distribution of the error was determined to use as boundary conditions for random error generation when testing the metabolic syndrome algorithm.

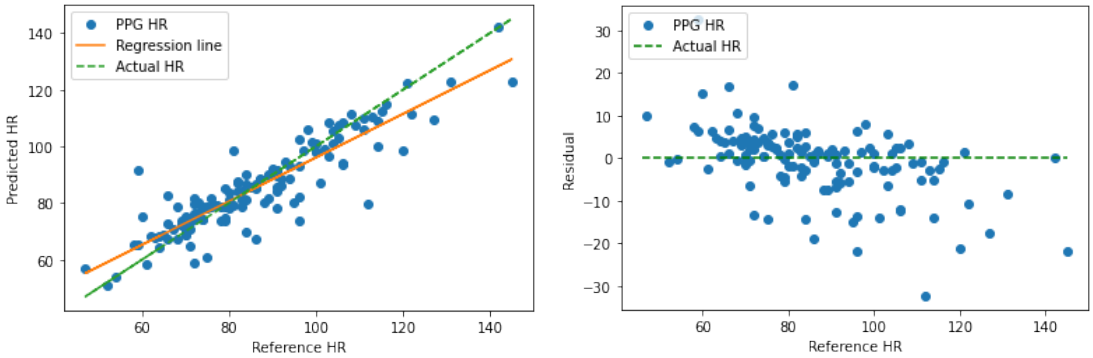
[1] Mean Error
[2] Mean Absolute Percent Error
[3] American National Standards Institute

The ME, MAE, and MAPE obtained from the 133 participants' smartphone PPG HR measurements are shown in Table 5.1.

Table 5.1: Results from the HR algorithm on smartphone PPG measurements

Parameter	**Result**
ME:	0.77 ± 5.58 BPM
MAE:	5.42 ± 4.22 BPM
MAPE:	6.46%

While the results shown in Table 5.1 are within the performance boundaries of 10% MAPE, the algorithm's performance still falls somewhat short of that achieved by the likes of MIT's Cardiio application, which achieved a MAPE of 1.40% for resting HR, 2.53% for postmoderate-intensity exercise and 1.93% for postvigorous-intensity exercise on its fingertip PPG-estimated HR [21]. Based on the meta-analysis by De Ridder that looked at the accuracy of different smartphone based HR monitoring applications [35], it seems that our performance is in line with what could be expected from the average HR application currently available.



(a) Regression plot of Application HR vs. Actual/ Reference HR

(b) Residual plot of Application HR vs. Actual/ Reference HR

Figure 5.1: Results from the HR algorithm on smartphone PPG measurements

A more detailed understanding of the performance can be derived from Figure 5.1. Figure 5.1a is a regression plot, showing the scatter plot data of the smartphone PPG data against the actual reference HR, a regression line for the smartphone PPG data and a reference line. From this curve, it can be seen that the HR measurements generally have good precision, except for a few outliers. This may be seen more clearly on Figure 5.1b, which is a residual plot, showing the error of each PPG HR measurement. From this plot, it can be seen that the majority of measurements have an error of less than 5 BPM, with several outliers having errors far over 10 BPM.

The reason for large errors appearing in some cases are due to the peak detection algorithm used, and the fine tuning required when assessing which peaks to keep and which to reject, especially in noisy signals where the amplitudes are often shifting. Due to a host of different factors, like movement detection in the specific PPG algorithm used, movement by the participants, wind, light and other outside factors, there are a handful of signals

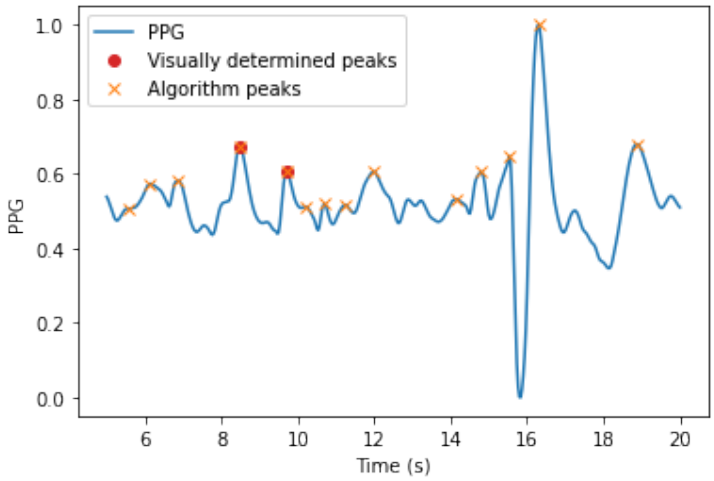


Figure 5.2: Example of a noisy signal with a large HR error

that were difficult to analyse, even manually. An example of such a signal is shown in Figure 5.2. The orange x's represent the peaks detected by the algorithm and based on a visual inspection, red dots were added to two peaks of two clean pulses that could be identified in the measurement. The correct HR in this example was 73 BPM, but was measured as 85 BPM by the algorithm and 72 BPM using visual inspection. These issues point to considerations that may need to be made in future studies, like better movement and signal quality detection or a more controlled testing environment.

5.2 Blood pressure

The criteria given by the ISO 81060-2 standard in Table 5.2 was used to assess the performance of the blood pressure algorithm. It was used initially to assess the performance of the algorithm on test data from the MIMIC III dataset. For an indication of the real world results the data measured by the smartphone application was used. A residual plot for the mean error was used to verify that the error remained relatively constant across all blood pressures and that the measurement was not biased in any direction. The distribution of the errors was also determined, which was used as boundary conditions for random error generation when testing the MetS algorithm in Section 5.3.3.

Table 5.2: ISO 81060-2 specification for blood pressure testing [48]

Parameter	**Requirement**
Number of participants:	85
Reference measurement:	Any sphygmomanometer, with maximum error $\leq \pm 1$ mm Hg
Pass criteria or value:	Mean error $\leq \pm 5$ mmHg, standard deviation within 8 mmHg
	MAE ≤ 5 (Class A), 10 (B) and 15 (C) mmHg

5.2.1 MIMIC Testing Data

For training the BP model, the MIMIC III dataset used was divided into a training, validation and testing set. The model achieved an accuracy of 94.8% on the training set, 93.4% on validation and 93.9% on testing. The results in terms of mean error and mean absolute error from estimating blood pressure on the test dataset, can be seen in Table 5.3.

Table 5.3: Results from the BP algorithm on on the MIMIC III test dataset

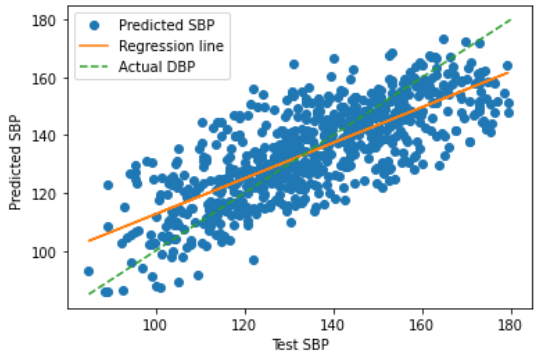
Parameter	**SBP**	**DBP**
ME:	-0.9 ± 13.2 mmHg	-0.4 ± 6.7 mmHg
MAE:	10.1 ± 8.5 mmHg	5.3 ± 4.2 mmHg

If compared to the criteria given by the ISO 81060-2 specification [48] for blood pressure testing in Table 5.2, the SBP results narrowly miss out on the criteria while the DBP passes with a classification of B, narrowly missing class A. While the mean error of the SBP is far below the maximum requirement of 5 mmHg, the standard deviation on the error is 13.2 mmHg, which is more than the required 8 mmHg. Based on the MAE of the SBP, if it were within the base requirement it would have been class C, narrowly missing class B due to the MAE of 10.1 mmHg being higher than the required 10 mmHg for class B.

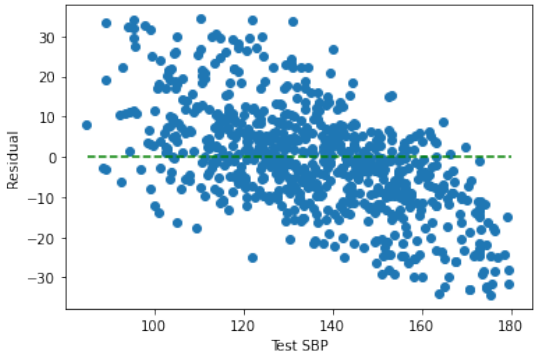
When looking at figures 5.3a and 5.3c, one would be forgiven for thinking that the SBP data actually performed better than the DBP data. In some aspects it did, such as the gradient of the regression line, which is closer to ideal . Something that is immediately apparent in Figure 5.3c is a lack of data across the range of values, which may have helped to lead to the better performance and may have also been one of the reasons for some of the issues that will be discussed later, in Section 5.2.2.

Looking at the regression lines of both Figures 5.3a and 5.3c, a clear downward gradient can be seen. This is just as apparent in the residual plots in Figures 5.3b and 5.3d. This indicates a tendency from the model to favour values around the mean, which could be due to an over-representation in training or simply a lack of features strong enough to provide greater variation. None of the variations made on the model, such as different preprocessing, network topology or optimisers, provided a better result in this regard.

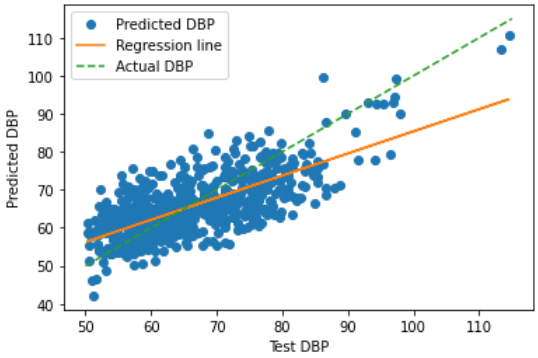
Despite the problems with the model that were pointed out, it still appears to be accurate enough in most cases to provide a good estimation for blood pressure. As can be seen in Figure 5.4, the majority of readings in both cases have errors less than 10 mmHg. For SBP in this set 35% of predictions were within 5 mmHg, 60% were within 10 mmHg and 77% were within 15mmHg. For DBP 55% were within 5 mmHg, 85% were within 10 mmHg and 97% within 15mmHg.



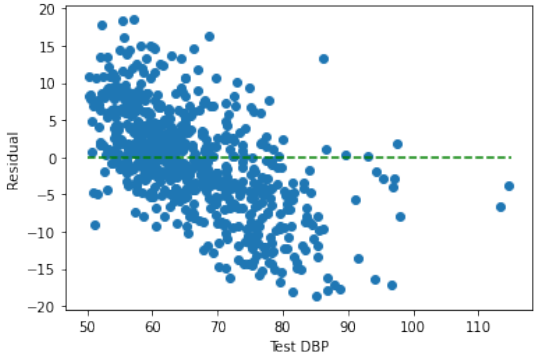
(a) Regression plot of smartphone PPG SBP vs. Actual Test SBP



(b) Residual plot of smartphone PPG SBP vs. Actual Test SBP



(c) Regression plot of smartphone PPG SBP vs. Actual Test DBP



(d) Residual plot of smartphone PPG SBP vs. Actual Test DBP

Figure 5.3: Results from the BP algorithm on MIMIC III test dataset

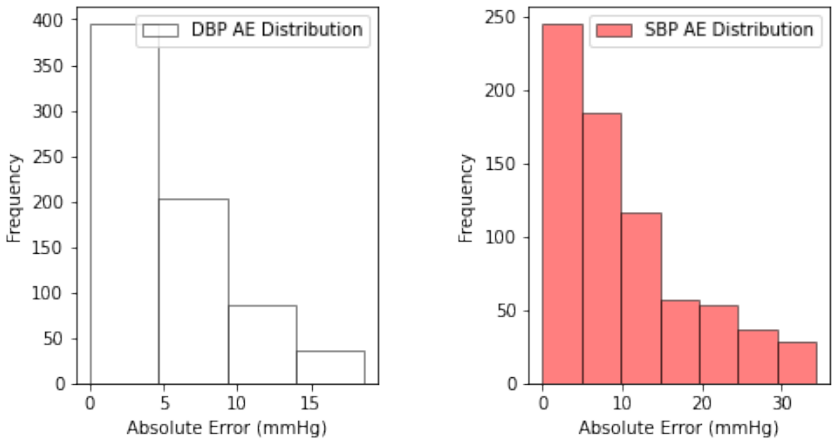


Figure 5.4: Distribution of absolute error of BP estimation on MIMIC PPG data

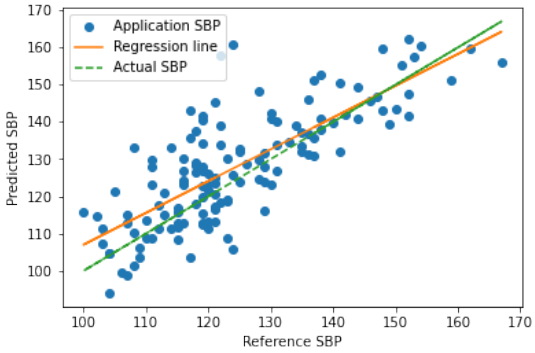
5.2.2 Smartphone Data

The results of estimating blood pressure from participants' smartphone PPG measurements can be seen in Table 5.4.

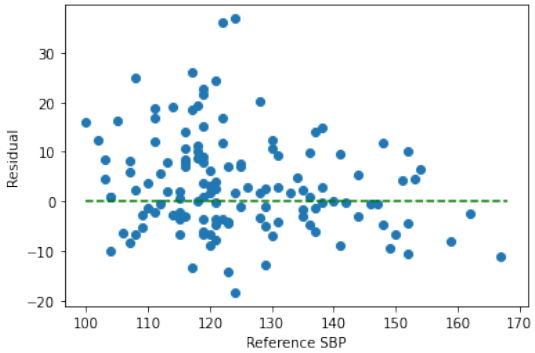
Table 5.4: Results from the BP algorithm on smartphone PPG measurements

Parameter	**SBP**	**DBP**
ME:	3.4 ± 9.8 mmHg	-5.2 ± 8.8 mmHg
MAE:	7.7 ± 6.9 mmHg	8.0 ± 6.3 mmHg

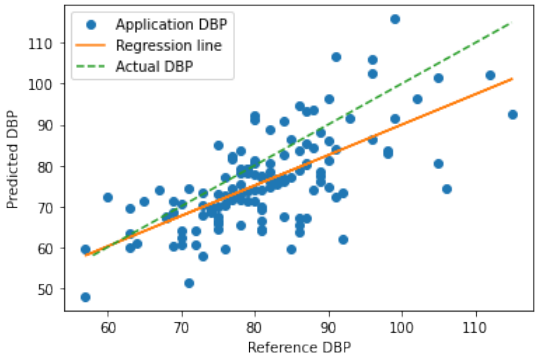
An unexpected anomaly that occurred in this case was that the smartphone PPG measurements outperformed the MIMIC data on SBP accuracy, however as could be expected, the DBP underperformed somewhat. If we also compare it to the ISO 81060-2 specification criteria, the SBP results still narrowly miss out on the criteria, although the standard deviation on the error is somewhat closer. The DBP narrowly fails this time, largely due to a big offset in the ME which also influenced standard deviation somewhat. Based on the MAE, if they were within the base requirements for ME, both the SBP and DBP would have been class B, as both fall between 5 and 10 mmHg.



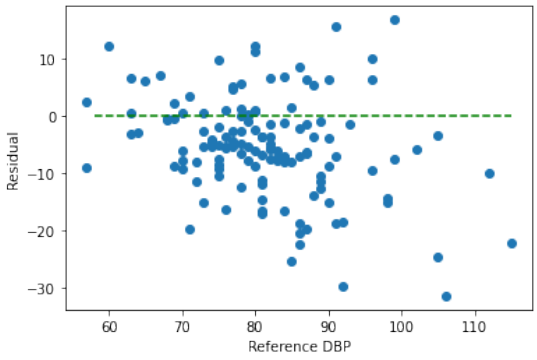
(a) Regression plot of MIMIC PPG SBP vs. Actual Test SBP



(b) Residual plot of MIMIC PPG SBP vs. Actual Test SBP



(c) Regression plot of MIMIC PPG SBP vs. Actual Test DBP



(d) Residual plot of MIMIC PPG SBP vs. Actual Test DBP

Figure 5.5: Results from the BP algorithm on smartphone PPG data

When looking at figures 5.5a and 5.5c, many of the same trends continue as was seen for the MIMIC data. One difference that is immediately apparent is a slight offset, upward for SBP and a slightly larger downward offset for DBP. Both cases could be a result of the tendency towards the mean that was previously noted.

Furthermore, in this case the DBP also tends to stay within a limited range, however not nearly as much as seen in the MIMIC data, and the average BP in this case seems to be slightly higher. This could be due to the target population of the MIMIC III dataset being ICU patients. It would be possible to shift the values up by the mean error, but it would be impossible to say if this tendency will continue for other datasets and on other devices. For the sake of transparency and keeping results as realistic as possible, the offset was left as is.

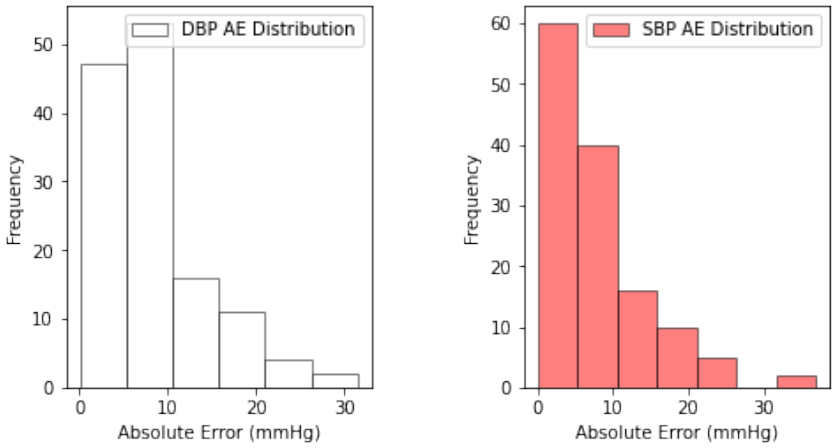
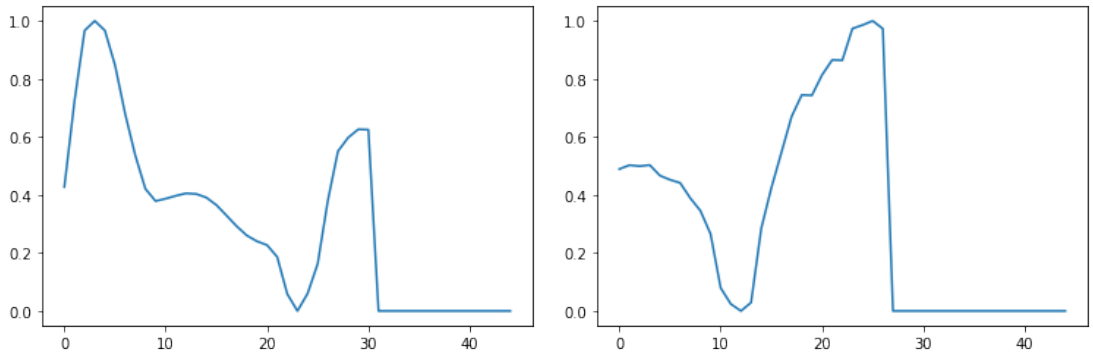


Figure 5.6: Distribution of absolute error of BP estimation on smartphone PPG data

In Figure 5.6, the offset for DBP is apparent again, causing the majority of it's readings to have a errors between 5 and 10 mmHg. However, despite this, by far the majority of predictions for both DBP and SBP have errors less than 10 mmHg.



(a) Single PPG wave from BP model input, with high error on DBP

(b) Single PPG wave from BP model input, with high error on SBP

Figure 5.7: Two examples of PPG portion of BP model inputs, with large errors

For the cases that had large errors, some investigation was done to see if a pattern could be recognised for what types of inputs would result in greater errors. Two examples of the time based PPG portion of the inputs (i.e. first 45 samples) can be seen in Figure 5.7, both of which had an error greater than 30 mmHg. In Figure 5.7a the error occurred for DBP and in the case of Figure 5.7b it was SBP. When looking at the waveform in Figure

5.7a, no glaring problem can be seen, except for a slightly slower second peak. Overall the signal quality looks good and if visually inspected would also likely have been chosen for measurement. On the other hand, the waveform in Figure 5.7b is clearly incorrect. This waveform was the result of one of the peak detection issues discussed in Section 5.1. Due to the nature of ANNs it is difficult to know exactly what features of the curves caused the errors. While the error in one would have been expected, the other is a mystery. Currently, any reasons given for these errors will be no more than speculation.

5.3 Metabolic Syndrome

For all the MetS models, training and testing data came from the SAPBA dataset [46]. Due to the limited size of the database (402 participants), to increase the reliability of test results and avoid biases, k-fold cross validation was used when assessing the model. The k variable was chosen as 10, due to the favourable results this value has shown in past work.

The procedure that was followed for the 10-fold cross validation was as follows:

1. The dataset was shuffled randomly
2. The dataset was split into 10 groups
3. For each of the 10 unique group:
   (a) The group was assigned as a hold out/ test data set
   (b) The remaining groups were combined and assigned as a training data set
   (c) The model was fitted on the training set and evaluated on the test set
   (d) The evaluation score was saved and the model discarded
4. The average scores across the 10 cycles were analysed to determine the skill of the classification technique.

In order to ensure a statistically relevant accuracy comparison of the ANNs, an approach similar to bootstrapping was followed where the 10-fold cross validation process was executed 10 times, randomly shuffling the data with each cycle, to result in a total of 100 training sets.

This process was repeated for both the initial MetS model and the MetS model including lifestyle factors. A slight variation of the process was followed when assessing the model with simulated BP and HR errors, where the training group remained standard but the test data received a random error based on the BP and HR results.

5.3.1 Without lifestyle factors

The results of the 10-fold cross validation, done on the MetS models using data of 402 participants from the SAPBA dataset [46], is given in Table 5.5. The results were analysed with several different performance indicators, each represented in a column in Table 5.5. For all these performance indicators, positive cases refer to a positive MetS diagnosis, and negative refers to no MetS.

The variable that is most often associated with ANN performance is accuracy. This was also the variable, together with loss, that all the MetS models were optimised on. Accuracy is the fraction of predictions that had the correct outcome. Accuracy is usually the most intuitive performance indicator and easiest to understand, however, can sometimes hide some biases which other indicators would be needed to detect.

The next column in Table 5.5 is precision, which simply put is the ability of the classifier not to label a sample that is negative as positive. Next is the recall score, which tests the other end of the spectrum than precision and is the ability of the classifier to find all the positive samples.

The last column is the F1 score, which can be interpreted as a harmonic mean of the precision and recall. The F1 score is one of the least intuitive performance indicators, but is valuable in that it is not as susceptible to biases in type-I and type-II errors as accuracy, and can often be a useful single indicator of performance.

Table 5.5: Cross validation results of the MetS algorithm, **without** lifestyle factors on 10 folds of the SAPBA dataset

	Test Samples	**Accuracy**	**Precision**	**Recall**	**F1 score**
Fold1:	41	0.634	0.538	0.824	0.651
Fold2:	41	0.659	0.625	0.556	0.588
Fold3:	40	0.900	0.917	0.917	0.917
Fold4:	40	0.825	0.882	0.750	0.811
Fold5:	40	0.750	0.667	0.750	0.706
Fold6:	40	0.650	0.600	0.375	0.462
Fold7:	40	0.825	0.857	0.818	0.837
Fold8:	40	0.700	0.647	0.647	0.647
Fold9:	40	0.675	0.647	0.611	0.629
Fold10:	40	0.900	0.867	0.867	0.867
Mean:		0.752	0.725	0.711	0.711

As can be seen in Table 5.5, over the 10 folds the model without the use of lifestyle factors, obtained an average accuracy of 75.2% and similar scores for all the performance indicators. Table 5.6 is a confusion matrix with the combined results from all 100 testing sets from bootstrapping, and it echoes the consistent performance with a very even split of false negative and false positive errors.

The average accuracy across the 100 sets was 73.4%, with a sensitivity of 73.1% and a

Table 5.6: Average confusion matrix of the MetS algorithm, **without** lifestyle factors across 100 training and testing cycles

		Predicted	
		Positive	**Negative**
Actual	**Positive**	33.30 ± 8.90 %	12.24 ± 5.70 %
	Negative	14.33 ± 7.31 %	40.13 ± 5.57 %

specificity of 73.6%. This means that the performance is lower than would be required to use it as a diagnostic tool, as it is somewhat weaker compared to different MetS definitions, IDF (Acc: 95.5%; Sen: 90.2%; Sp: 100%) and WHO (Acc: 87.4%; Sen: 78.2%; Sp: 95%). However, the value of the solution does not end with definite yes or no answers. The output from the ANN was a single digit with a value ranging from 0 to 1. To determine what the network's prediction was, this output was rounded and the resulting 0 or 1 was used to determine the MetS prediction (i.e. 0 = Negative, 1 = Positive). The output value can also, in a sense, be interpreted as a confidence value with 0 indicating full certainty in a negative diagnosis, 0.5 being complete uncertainty and 1 being full certainty in a positive diagnosis.

To test the viability in interpreting this output as a risk indicator instead of a diagnosis, the accuracy was calculated for output values with a confidence above 80% (that is output values below 0.2 or above 0.8) and 90% (outputs below 0.1 or above 0.9).

Table 5.7: Cross validation accuracy of the MetS algorithm, **without** lifestyle factors for only cases with 80+% and 90+% confidence

	80+% Confidence		**90+% Confidence**	
	Test Samples	Accuracy	Test Samples	Accuracy
Fold1:	18	0.611	7	0.714
Fold2:	17	0.824	7	1.000
Fold3:	18	0.944	3	1.000
Fold4:	21	0.857	7	1.000
Fold5:	16	1.000	12	1.000
Fold6:	14	0.857	11	0.909
Fold7:	19	0.895	11	0.909
Fold8:	17	0.941	11	0.909
Fold9:	13	1.000	6	1.000
Fold10:	20	0.950	10	1.000
	Total:	**Weighted mean:**	**Total:**	**Weighted mean:**
	173	0.884	85	0.941

As can be seen from the results in Table 5.7, there is a clear increase in performance as the confidence level of the output increases. As can also be seen from this table, there is still a significant proportion of sample size (173/402 samples, 43%) that fall within the 80% confidence levels. This seems to indicate that there is definite value in using this model as a risk indicator, which a user could potentially use to determine whether a full

diagnosis is warranted. In that sense, it falls perfectly in line with what is expected from a screening solution.

5.3.2 With lifestyle factors

The results of the 10-fold cross validation, done on the MetS models using data including lifestyle factors from the SABPA dataset [46], is given in Table 5.5. As can be seen from the results in Table 5.5, there was a very slight improvement in performance across all areas compared to the model excluding lifestyle factors.

Table 5.8: Cross validation results of the MetS algorithm, **with** lifestyle factors on 10 folds of the SAPBA dataset

	Test Samples	**Accuracy**	**Precision**	**Recall**	**F1 score**
Fold1:	41	0.659	0.565	0.765	0.650
Fold2:	41	0.732	0.667	0.778	0.718
Fold3:	40	0.850	0.950	0.792	0.864
Fold4:	40	0.800	0.875	0.700	0.778
Fold5:	40	0.725	0.632	0.750	0.686
Fold6:	40	0.750	0.714	0.625	0.667
Fold7:	40	0.825	0.857	0.818	0.837
Fold8:	40	0.775	0.750	0.706	0.727
Fold9:	40	0.725	0.733	0.611	0.667
Fold10:	40	0.875	0.812	0.867	0.839
Mean:		0.772	0.756	0.741	0.743

The increased accuracy across the board is also made evident by the confusion matrix in Table 5.9, as there was a decrease in both type-I and type-II errors. The average accuracy across the 100 sets was 76.3%, with a sensitivity of 74.9% and a specificity of 77.4%. As seen by the relation between the number of false positives and false negatives in Table 5.9, as well as the precision, recall and F1 score in Table 5.8 this model is also marginally less biased. A t-test was done in order to ensure that the increase in performance was not just an anomaly, and the accuracy difference across the 100 training and testing cycles was found to be statistically relevant ($t = -2.48$; $p < 0.014$).

Table 5.9: Average confusion matrix of the MetS algorithm, **with** lifestyle factors across 100 training and testing cycles

		Predicted	
		Positive	**Negative**
Actual	**Positive**	33.77 ± 8.64 %	11.77 ± 5.02 %
	Negative	12.47 ± 6.86 %	41.99 ± 5.71 %

One area, however, where a performance increase was not seen was when looking at higher confidence levels, as seen in Table 5.10. The decreased accuracy compared to the model

excluding lifestyle factors may also be due to a higher number of samples included in the higher confidence levels. There were 182 test samples included in this 80% confidence set, compared to the previous 173 and 96 in the 90% group, compared to 85. Nevertheless, this is largely irrelevant because as with the previous model, an overall increase in accuracy is still evident when looking at the higher confidence levels, indicating the screening method proposed remains viable for this model as well.

Table 5.10: Cross validation accuracy of the MetS algorithm, **with** lifestyle factors for only cases with 80+% and 90+% confidence

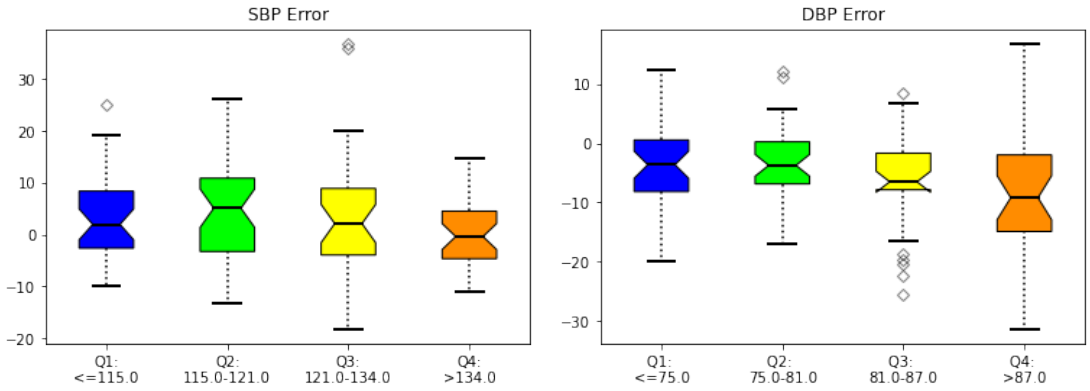
	80+% Confidence		**90+% Confidence**	
	Test Samples	Accuracy	Test Samples	Accuracy
Fold1:	18	0.722	8	0.750
Fold2:	17	0.706	11	0.909
Fold3:	16	1.000	8	1.000
Fold4:	17	0.882	6	1.000
Fold5:	24	0.917	15	0.933
Fold6:	16	0.812	9	1.000
Fold7:	19	0.842	10	0.900
Fold8:	17	0.941	10	0.900
Fold9:	19	0.842	10	0.900
Fold10:	19	1.000	9	1.000
	Total:	**Weighted mean:**	**Total:**	**Weighted mean:**
	182	0.868	96	0.927

5.3.3 Application data error simulation

In order to simulate real-world scenarios random errors were generated on heart rate and blood pressure based on the results from the smartphone data. Each dataset (SBP, DBP and HR) determined during the data collection phase was divided into four parts, based on the values measured during testing. The data was divided at the lower quartile, median and upper quartile. The distribution of these errors in the form of box plots can be seen in Figures 5.8a for SBP and 5.8b for DBP.

The *random.choices* function from the Python Standard Library [77] was used to generate random errors on heart rate and blood pressure based on the errors from Figure 5.8. For each section, all errors were rounded to the nearest integer, tallied and normalised to determine the weights to be used for the *random.choices* function to simulate real world circumstances.

In order to get an apples to apples comparison, the test predictions with errors were done in parallel to MetS predictions with lifestyle factors. This way the results received were from the exact same models, trained on the exact same data, but with an error added to the test data. We therefore know that any deviations from the results seen in Table 5.8 is wholly due to the added BP errors.



(a) Box plot of SBP errors per quartile (b) Box plot of DBP errors per quartile

Figure 5.8: Box plots for (a) SBP and (b) DBP PPG-measurement estimation errors per quartile

Table 5.11: Cross validation results of the MetS algorithm with BP error simulation

	Test Samples	**Accuracy**	**Precision**	**Recall**	**F1 score**
Fold1:	41	0.610	0.529	0.529	0.529
Fold2:	41	0.683	0.647	0.611	0.629
Fold3:	40	0.825	0.947	0.750	0.837
Fold4:	40	0.725	0.846	0.550	0.667
Fold5:	40	0.750	0.714	0.625	0.667
Fold6:	40	0.725	0.667	0.625	0.645
Fold7:	40	0.775	0.842	0.727	0.780
Fold8:	40	0.750	0.733	0.647	0.688
Fold9:	40	0.700	0.750	0.500	0.600
Fold10:	40	0.950	1.000	0.867	0.929
Mean:		0.749	0.768	0.643	0.697
Reference:		0.772	0.756	0.741	0.743

The results of the 10-fold cross validation done on the MetS model including lifestyle habits and including an error added to the SBP, DBP and HR, is given in Table 5.11. On first glance, when looking at the accuracy, the results in Table 5.11 indicate a very slight drop in performance from the results without the BP en HR errors. However, when looking at the confusion matrix in Table 5.12, it is evident that there is a significant increase in false negative errors (from 11.4% in Table 5.9 to 15.75% in 5.12). The accuracy was not effected severely by the noted bias because there was also a decrease in false positive errors, which indicates a bias towards negative diagnoses. This result can also be seen in the recall, which is dependent on false negative errors, and F1 scores in Table 5.11, as it is sensitive to biases. This bias in the MetS results is very likely due to the negative bias noted in the experimental DBP results earlier. This clearly illustrates a very high reliance on the BP measurements by the MetS algorithm.

Based on Table 5.12, the average accuracy across the 100 testing sets was 73.1%, with a sensitivity of 64.9% and a specificity of 79.9%. These results reaffirm the bias that was previously noted.

Table 5.12: Average confusion matrix of the MetS algorithm, **with** lifestyle factors and added errors across 100 training and testing cycles

		Predicted	
		Positive	**Negative**
Actual	**Positive**	29.57 ± 8.38 %	15.97 ± 6.00 %
	Negative	10.96 ± 5.80 %	43.50 ± 6.17 %

5.4 Usability

5.4.1 Effectiveness

As described in the introductory chapter, effectiveness was determined using the following formula:

$$Effectiveness = \frac{Number\ of\ tasks\ completed\ successfully}{Total\ number\ of\ tasks\ undertaken} \times 100\% \tag{5.1}$$

During the data collection procedure, participants only had a single task to complete, i.e. take a single PPG measurement, which means that for most cases the top of the fraction in equation 5.1 would be equal to one. However, there were several participants who opted to take another measurement. In these cases the number of recorded measurements were counted and the number of attempts for each measurement were summed.

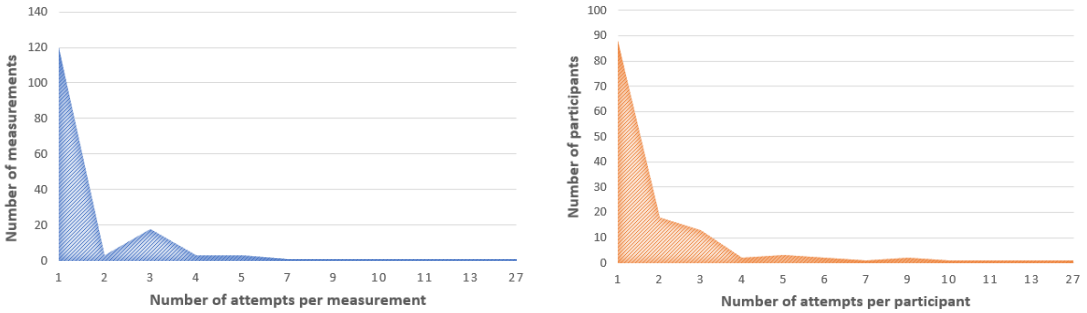
From the 133 participants who took part in the data collection, there were 153 successful PPG measurements and 284 total measurement attempts.

As summed up in Table 5.13, 114 (85.7%) participants only took a single measurement, 18 (13.5%) took two measurements and only a single participant opted for a third.

Table 5.13: Number of participants to number of measurements taken

Number of measurements	**Number of participants**
1	114
2	18
3	1

When calculating the effectiveness with which users were able to complete their measurements according to equation 5.1, the majority (77%) of participants had a score of 100%. The exact score tallies can be seen in Table 5.14 with a graph displaying the same results in Figure 5.10.



(a) Number of measurements made (y) vs. the number of attempts (x) required for measurement

(b) Number of participants (y) vs. the number of attempts (x) made by each participant

Figure 5.9: Tally of measurements (a) and participants (b) based on number of attempts required

Table 5.14: Number of participants to obtain different levels of effectiveness

Effectiveness (%)	**Number of participants**
100	103
50	4
33	16
25	1
20	3
14	1
11	1
10	1
9	1
8	1
4	1

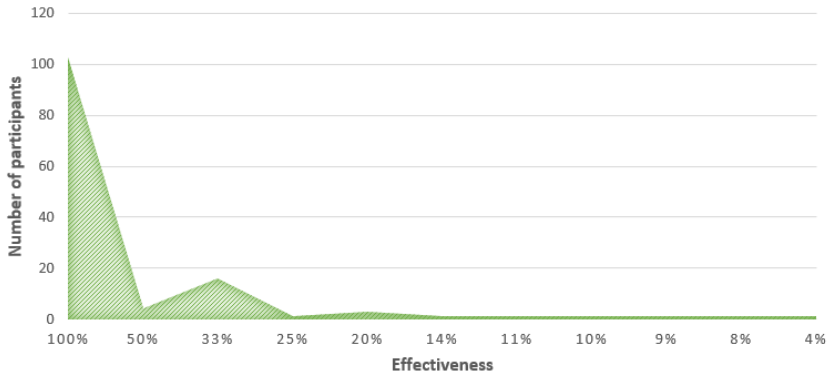


Figure 5.10: Number of participants to obtain different levels of effectiveness

The average effectiveness was calculated to be 84% with a minimum of 4%, maximum of 100% and a standard deviation of 30%.

5.4.2 Efficiency

Time based efficiency would usually be determined as follows:

$$\overline{P}_t = \frac{1}{NR} \sum_{j=1}^{R} \sum_{i=1}^{N} \frac{n_{ij}}{t_{ij}} \tag{5.2}$$

where N is the total number of tasks, R is the number of users, n_{ij} is the result of task i by user j ($n_{ij} = 1$ for successful task and 0 for unsuccessful) and t_{ij} is the time spent completing the task. The result has a unit of tasks/sec, meaning it is effectively the inverse of the average time taken for successful tasks. This provides little value as a performance indicator, since there are no reference systems for it to be compared to. However, since the calibration time is standardised to 5 seconds and the measuring time 15 seconds, we can calculate an ideal efficiency to be:

$$\overline{P}_t(ideal) = \frac{1}{t_{calibrate} + t_{measure}} \tag{5.3}$$

$$\overline{P}_t(ideal) = \frac{1}{20} \tag{5.4}$$

$$\overline{P}_t(ideal) = 0.05\,\text{tasks/s} \tag{5.5}$$

A more valuable metric that was calculated was relative efficiency, which is calculated as follows:

$$\overline{P} = \frac{\sum_{j=1}^{R} \sum_{i=1}^{N} n_{ij} t_{ij}}{\sum_{j=1}^{R} \sum_{i=1}^{N} t_{ij}} \times 100\% \tag{5.6}$$

By also using the time of the tasks that were not completed, we get a percentage to represent effectiveness. However, equation 5.6 was modified slightly to better fit the available data. Due to the way that the PPG measurement works there is no clear way to determine exactly when a failed measurement attempt ends and a new one begins, though we can use the ideal time of 20 seconds for a measurement as a standard for successful measurements. This means that equation 5.6 will change to:

$$\overline{P} = \frac{20R}{\sum_{j=1}^{R} T_j} \times 100\% \tag{5.7}$$

where T is the total time taken by the participant to get a successful measurement.

The time taken to complete each of the 153 PPG measurements, ranged from 26 to 226 seconds. The average time taken to successfully complete a measurement was 45.7 seconds, which means that, considering an ideal time of 20 seconds, on average participants took

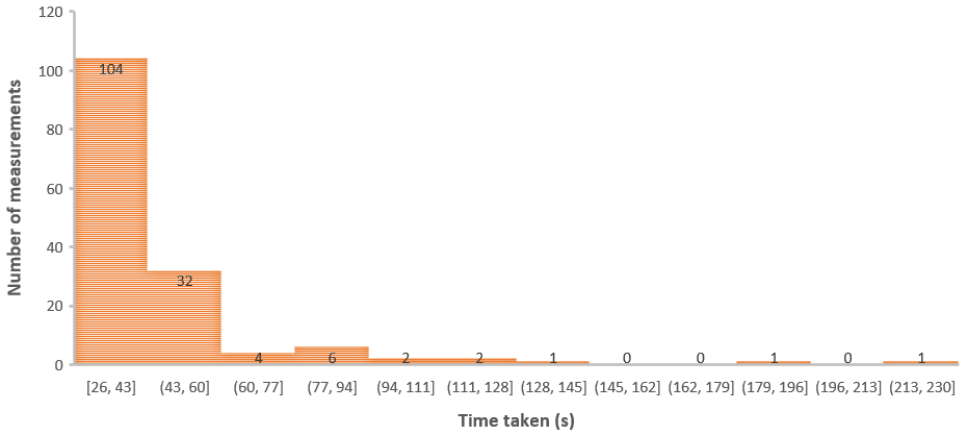


Figure 5.11: Distribution of time taken to complete a successful measurement

just as long to read the instructions, position their fingers and restart measurements as they did taking the successful PPG measurement.

Figure 5.11 shows a more detailed overview of the time taken to take a successful PPG measurement. From the figure it can be seen that 104 of the 153 measurements (68%) took less than 43 seconds to complete. The majority of cases that took longer were those that had trouble taking a successful measurement in a single attempt.

To get a better idea of what the efficiency is for single measurements, we used equation 5.2, including only the measurements taken with just a single attempt.

$$\overline{P}_t = \frac{1}{120}\sum_{j=1}^{120}\frac{1}{t_j} \tag{5.8}$$

$$\overline{P}_t = \frac{3.33}{120} \tag{5.9}$$

$$\overline{P}_t = 0.028\,\text{tasks/s} \tag{5.10}$$

This shows an efficiency almost half of the ideal, of 0.05 tasks/s. However, this only paints half the picture, as unsuccessful attempts had to be excluded. A far more realistic metric was obtained when using equation 5.7.

$$\overline{P} = \frac{20 \times 153}{6988.48} \times 100 \tag{5.11}$$

$$\overline{P} = 44\,\% \tag{5.12}$$

This further reinforces the statement from the start of this section that on average less than half of the time used is on the actual measurement. However, even with the added time it still remains within the realm of ordinary blood pressure measurements and may even be quicker and easier once users are accustomed to the measurement.

5.4.3 Satisfaction

To determine user satisfaction, participants were asked to provide four answers using a rating scale between 0 and 5 in increments of 0.5 based on how much they agreed or disagreed with a statement. The statements were as follows:

1. The app was straight forward to use.
2. I had very little trouble taking my blood pressure measurement.
3. I feel that I would require technical assistance to use this app.
4. If it were available, I would personally use an app like this.

Of the 133 participants, 129 provided feedback for the usability of the application. The average score for each statement can be seen in Table 5.15.

Table 5.15: Usability ratings provided by users for PPG measurement

	Statement	**Average Score**
1	The app was straight forward to use.	4.91
2	I had very little trouble taking my blood pressure measurement.	4.79
3	I feel that I would require technical assistance to use this app.	1.11
4	If it were available, I would personally use an app like this.	4.71

The scores were all rescaled to give a score out of 100. As a score of 0 would be the ideal in that scenario, the third statement's score was inverted so that 100 is the best score. The scores were assessed with reference to the SUS provided in Figure 1.5, which can be seen in Table 5.16.

Table 5.16: Rescaled usability ratings with SUS Scores

	SUS Score	**SUS Grade**	**Rating**
1	98.2	A	Excellent
2	95.8	A	Excellent
3	77.8	C	Good
4	94.2	A	Excellent

The majority of users provided the maximum score to statement one and two, indicating that the application and the measurement process was straight forward and intuitive. There was a very slight dip for the fourth statement, which is understandable, considering that many users would not have much of a personal interest in knowing their blood pressure and would therefore select 4/5 or less. This still seems to only apply to the minority of participants. The point with the lowest score was statement three, which may be because some participants did not find the diagrams and instructions in the application adequate and still somewhat relied on the explanation from the researcher.

Another possibility is that the layout and wording of the question may have caused some confusion, as this was the only statement where 0 would have provided the best score.

Nevertheless, the average score across the four statements is 91.5, which is an excellent rating and an A on the grade scale.

5.5 Chapter summary

In this chapter the results obtained from the experimental procedure were discussed. The results are divided according to heart rate, blood pressure, MetS and usability.

The standard that was used to asses the algorithm to obtain heart rate from PPG measurements, was the specification given by the AAMI, the Consumer Technology Association [74] and ANSI [75], which required a MAPE rate of less than 10%. The HR algorithm used on experimentally obtained PPG data was found to have a ME of 0.77 ± 5.58 BPM, MAE of 5.42 ± 4.22 BPM, and MAPE of 6.46%. These values fall within the specified requirements, however the algorithm's performance still falls somewhat short of that achieved by similar applications [21], [35].

The accuracy of the blood pressure algorithm was determined on a test set from the MIMIC dataset and the experimentally acquired PPG data. In both cases, it was assessed according to the ISO 81060-2 specification for blood pressure testing [48].

The MIMIC test set obtained an an accuracy of 93.9%, with a ME of -0.9 ± 13.2 mmHg for SBP and -0.4 ± 6.7 mmHg for DBP, and a MAE of 10.1 ± 8.5 mmHg and 5.3 ± 4.2 mmHg for SBP and DBP respectively. The smartphone PPG data had a ME of 3.4 ± 9.8 mmHg for SBP and -5.2 ± 8.8 mmHg for DBP, and a MAE of 7.7 ± 6.9 mmHg and 8.0 ± 6.3 mmHg for SBP and DBP respectively. If we compare these results to the ISO 81060-2 specification criteria [48], in both cases the SBP results narrowly miss out on the base criteria. For the MIMIC set, the DBP passes with a classification of B, narrowly missing class A. However, the smartphone DBP narrowly fails, largely due to a large offset in the ME, which also influenced standard deviation somewhat.

The two different MetS classifiers were tested using 10-fold cross validation. The first network had an accuracy of 75.2%, precision of 72.5%, recall of 71.1% and an F1 score of 71.1%. The second, including lifestyle factors, had an accuracy of 77.2%, precision of 75.6%, recall of 74.1% and an F1 score of 74.3%. In order to ensure that the difference in performance noted between the two ANNs is statistically relevant, an approach similar to bootstrapping was followed where the 10-fold cross validation process was executed 10 times, randomly shuffling the data with each cycle, to result in a total of 100 sets. The average accuracy across the 100 sets for the first model was 73.4%, with a sensitivity of 73.1% and a specificity of 73.6%. For the model with lifestyle factors the accuracy was 76.3%, with a sensitivity of 74.9% and a specificity of 77.4%. A t-test was done to show that the accuracy difference between the models was statistically relevant ($t = -2.48$; $p < 0.014$).

By interpreting the output value of the ANN as a confidence level, with 1 being 100%

percent certainty of a positive diagnosis and 0 being 100% certainty of a negative diagnosis, we could derive more value from the output of the model. When looking at the accuracy of the models at only confidence levels above 80%, the two models achieved an accuracy of 88.4% and 86.8%. When looking at only confidence levels over 90%, the models achieved 94.1% and 92.7%. This indicates a definite improvement in accuracy as confidence levels increase.

A third set of cross validation was done on the exact same ANNs as the model including lifestyle factors, with an error added to the SBP, DBP and HR based on the experimentally obtained results. The results from this cross validation were an accuracy of 74.9%, precision of 76.8%, recall of 64.3% and an F1 score of 69.7%. The average accuracy across the 100 testing sets, from the bootstrapping method, was 73.1%, with a sensitivity of 64.9% and a specificity of 79.9%. These results show that the negative bias from the DBP greatly influenced the MetS model, which is to be expected considering the observations made by [65].

With regards to usability, effectiveness was calculated by getting the ratio of the number of successful measurements a user made and the total measurement attempts. The majority (77%) of participants obtained a score of 100%, therefore only needing a single attempt to take a measurement. The average effectiveness was calculated to be 84% with a minimum of 4%, maximum of 100% and a standard deviation of 30%.

Efficiency was determined as the total time spent on taking successful measurements divided by the total time taken. An overall efficiency of 44% was achieved. Which, while it may seem low, is standard if not better compared to standard cuff-based blood pressure monitors. It is also important to keep in mind that this was most users' first time taking a PPG measurement, and efficiency would increase once users are more proficient in the process.

To determine user satisfaction, participants were asked to provide four answers using a rating scale between 0 and 5 in increments of 0.5 based on how much they agreed or disagreed with a statement. Of the 133 participants, 129 provided feedback for the usability of the application. The scores were rescaled to give a score out of 100, and the average score across the four statements was 91.5%, which, based on the SUS in Figure 1.5 is an excellent rating and an A on the grade scale.

CHAPTER

6

CONCLUSION

Metabolic syndrome is reaching epidemic status, contributing to some of the biggest causes of death nationwide and worldwide. MetS contributes to some major diseases such as heart disease, coronary artery disease and type 2 diabetes. Since MetS is, in most cases, easily preventable and reversible with lifestyle changes, if people were better informed about their status many of these conditions and deaths could possibly be prevented.

In this dissertation we developed a method of screening for MetS using a mobile application, in the hope that an application of this kind could prove to be a valuable tool for informing the public of their health condition. This application made use of user entered data as well as PPG readings obtained using a cell phone camera. This information could then be used by an ANN to determine the likelihood of the user having MetS.

The question posed in the problem was: "How accurately could a smartphone screen for metabolic syndrome, and would this be a feasible solution?" The question was posed in such a way that sufficiently answering it would require verification of the accuracy of the screening method as well as the validation that it is a feasible solution to the problem.

With the problem in mind, considerations had to be made about what would facilitate a successful screening application. Herman and Gill, define the ideal medical screening test as one that is accessible, simple to use, inexpensive, and associated with minimal discomfort to the person/s being screened, however, they did not provide an ideal or threshold accuracy to be reached [31]. We could start to assess the application by the terms set in [31]:

- **Accessible:** Due to the popularity of smartphones, they have been a major topic of study for medical devices in recent years. Due to all the research in this area,

a new field, known as mHealth, was formed [32]. Their accessibility is the driving factor behind creating smartphone based screening solutions.

- **Inexpensive:** In order to keep the application as inexpensive as possible, and thereby also increasing the accessibility, a requirement was set not to include any external sensors. The application only made use of features and sensors that would be available on a standard smartphone.
- **Simple to use:** This was analysed according to the ISO 9241-11 specification by determining users' effectiveness, efficiency and satisfaction on the application.
- **Associated with minimal discomfort:** This falls much in line with the previous point, but was also an important point to take into account when considering the type of tests to include. The features were limited to fields that a user must fill in and a PPG measurement, to avoid unnecessary discomfort or inconvenience that may be associated with things like calorie counting or even traditional cuff-based blood pressure measurements.

In terms of accuracy, there was no definite goalpost, as the accuracy required by a screening method tends to largely vary based on the specific application or method in question. The main requirement was that the application would provide a result accurate and detailed enough that the user would have a better understanding of their current health status with regards to MetS and some of its risk factors, which they could use to assess whether a diagnosis from a medical professional would be necessary. As a benchmark, different definitions were used to determine MetS on the SABPA dataset, which gave an indication of the type of variation that could be expected between different definitions of MetS and thereby know how close the proposed solution is to standard diagnostic tools. Furthermore, to ensure that the accuracy was maximised, each measurement step, up to the final classification, was assessed according to the standards relevant to that measurement.

Two of the inputs required for the MetS algorithm, HR and BP, needed to be determined using a PPG waveform, that could be measured using a smartphone camera. The PPG measurements were made by determining a threshold based on the minimum and maximum values from the red channel during a calibration phase and determining the number of red pixels over each frame in the measuring stage. HR was determined by measuring the peak-to-peak time over several beats in a full measurement. Outliers were then removed before calculating an average to determine HR

The BP model was trained using clinical PPG and BP data from the MIMIC III database, by resampling PPG signals down to 30 Hz to match the frame rate from most smartphone cameras, applying a bandpass filter to remove sampling noise and low frequency movements then isolating a single waveform. An FFT was determined of the waveform, which was then used together with the time domain wave to train an ANN to determine SBP and DBP.

Two different MetS classifiers were built, with slightly different input features. The first model only included user entered biometric information (sex, age, BMI, waist-height ratio)

and measured HR and BP. The second model included all the features from the first, including lifestyle factors (smoking, drinking, activity level) and medical history.

Table 6.1: Summary of real-world and error-simulated results across all variables

Variable	Dataset	Metric	Result
HR	Smartphone	MAPE	6.46 %
SBP	Smartphone	ME	3.4 ± 9.8 mmHg
		MAE	7.7 ± 6.9 mmHg
DBP	Smartphone	ME	-5.2 ± 8.8 mmHg
		MAE	8.0 ± 6.3 mmHg
MetS	SABPA with BP and HR errors from smartphone data	Accuracy	73.1 %
		Sensitivity	64.9 %
		Specificity	79.9 %
Usability	Smartphone	Effectiveness	84 %
		Efficiency	44 %
		Satisfaction	91.5 %

Table 6.1 provides a summary of the results achieved during the real world testing (or error simulation in the case of MetS) for all the test parameters.

The accuracy of the HR algorithm used on experimentally obtained PPG data falls within the specified requirement of 10%, however it is not necessarily a good indication of the performance that could be expected from the state of the art. From deeper investigation into individual errors, it was determined that the majority of measurements were much closer than would be suggested from the average result. There were a handful of outliers that, because of their large errors, substantially decreased the overall performance.

One reason for large errors appearing in some cases is due to the peak detection algorithm used, and the fine tuning required when assessing which peaks to keep and which to reject, especially in noisy signals where the amplitudes are often shifting. There were a host of different factors, like movement detection in the specific PPG algorithm used, movement by the participants, wind, light and other outside factors, that contributed to making some signals more difficult to analyse. In some of these cases, the signal quality was so poor that the HR could also not be determined by manual visual inspection.

The issues experienced in the HR research point to the following considerations that may need to be made in future studies:

- **Better movement and signal quality detection:** The method of movement and signal quality detection that was used was purely statistical, counting on the predictable conditions in PPG measurement to set measurement boundaries. However, based on the measurement results, these boundaries were not sufficient in accounting for all the possible signal distortions that may occur.

- **A more controlled testing environment:** A major drawback of the location where testing took place that was encountered was a lack of control of outside conditions such as wind and light. An indoor venue would have allowed better control of these variables, to provide better results and also eliminate them as factors when encountering measurement errors.
- **ECG reference for post-exercise HR measurement:** Because a sphygmomanometer was used for BP and HR measurement that meant that the two HR measurements could not happen simultaneously and in the time between measurements of post-exercise participants a drop in HR would often occur. For future studies if post-exercise HR is tested, it is advised to use ECG or similar methods that would allow multiple measurement types at a time.

None of the BP models met the full requirements of the ISO 81060-2 specification for blood pressure testing. However, in all cases the margins by which the specifications were missed were small enough that the results from the PPG based BP estimation still showed great promise. Performance dropped slightly with the change to a smartphone camera, compared to the MIMIC data, especially in the DBP readings. Adding an extra calibration step could possibly provide some, limited, improvement. However, this would only ensure improvement for the test dataset, and could possibly mean that a different device or different subset of data would see a drop in performance. This would also not be an ideal scenario as it would mean that standardisation across devices would not be possible.

Despite the inaccuracies encountered, cellphone based blood pressure measuring remains incredibly promising, and warrants further dedicated research into the field. Some of the following items can be considered for future research or development:

- **Dedicated smartphone PPG datasets:** Considering the amount of research currently going into the topic, the formation of a large dataset containing smartphone based PPG measurements, BP measurements, HR and other applicable parameters would be a venture that could be incredibly valuable to future research in the field.
- **Multi-device testing:** While the design approach used when developing the BP model in this dissertation accounted for the use on different types of smartphones or other forms of PPG measurement, physical testing was all done with a single device. Before an application using this model could be launched, sufficient testing would have to happen to ensure accuracy across a large range of devices.

As seen in Table 6.2, the results achieved by the MetS models deviates far more from the reference SABPA method than that seen by other standard MetS definitions. However, it still performs well enough to provide a reasonable indication and using the confidence level of the output means that users could use the added information to have a better understanding of what their results are. The added benefit of this route would be that, because a confidence level is included, users would be less likely to treat the results like a diagnosis, perhaps causing unwarranted concern or confidence regarding their health status.

Table 6.2: Summary of results from IDF and WHO MetS definitions and ANN models compared to SABPA method

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
IDF	95.5	90.2	100
WHO	87.4	78.2	95
ANN without lifestyle factors	73.4	73.1	73.6
ANN with lifestyle factors	76.3	74.9	77.4
ANN with BP and HR errors	73	64.9	79.9

There are four recommendations for future research into MetS classifiers:

- **SABPA dataset size:** A major limiting factor when training the MetS models was the size of the training set. There are other, larger datasets available that were not used in this dissertation because, once the need was realised, time did not allow for the extra data processing and sifting that would be required. It is recommended that future research make use of these datasets, or perform their own large scale data collection.
- **Multi-parameter output layer:** Upon further consideration, another option that was also considered for a MetS screening tool was to train a network on the number of positive risk factors, instead of just the MetS outcome. This could be done by scaling the output as a single analog value between 0 and 1. This may provide more accurate loss indications, and may also add value to the confidence based risk-indicator approach.
- **Deeper feature analysis:** At the end of this research there remained some questions regarding the impact of each feature on the MetS model. Several trial and error tests were done throughout the course of the research to verify appropriate features, and some features were also excluded to see if the dimension reduction would affect the performance. In most cases the proposed model performed best. However, there may be value in doing an in-depth analysis on the impact of every feature on MetS to develop a theoretically ideal model.
- **Regularisation to reduce overfitting:** While overfitting was noted and accounted for in the neural network design, the use of regularisation was not investigated. Future research could look into regularisation techniques such as dropout, early stopping, weight decay, etc., to improve performance while reducing overfitting.

Based on the ISO 9241-11 standard, usability of the application was measured in terms of effectiveness, efficiency and satisfaction.

The effectiveness with which users were able to use the application was calculated by getting the ratio of number of successful measurements a user made and the total measurement attempts. The majority (77%) of participants obtained a score of 100%, therefore

only needing a single attempt to take a measurement and 92.5% of participants needed three attempts or less.

Efficiency was determined as the total time spent on taking successful measurements divided by the total time taken. An overall efficiency of 44% was achieved. Which, while it may seem low, when taking into consideration that it was determined with respect to the ideal time of 20 seconds one would realise that it is in fact comparable if not better compared to cuff-based blood pressure monitors. It is also important to keep in mind that this was most users' first time taking a PPG measurement, meaning both efficiency and effectiveness would increase once users are more proficient in the process.

User satisfaction was determined based on responses that were received from participants to four statements. The average score across the four statements was 91.5%, which, based on the SUS in Figure 1.5 is an excellent rating and an A on the grade scale. This score may even be a slightly lower representation than participants' actual opinions, since one of the statements required responses in the opposite manner to the rest and it was noticed during the research that this caused some confusion for some participants.

A major limiting factor with regards to determining the usability of the application was that the data collected in the data gathering process only included PPG, HR and BP measurements. This was based on the capabilities of the researchers, and not being able to measure invasive risk factors for MetS. This meant that participants did not get a perspective of the entire application, but rather just a single aspect thereof, i.e. the PPG measurement. While this means that the results were only limited to a single element of the application, considering that PPG measurement is the most time consuming, involved and experimental aspect of the application, this should still provide a valuable indication of what the usability of a full MetS application would be.

To conclude we will take one final look at the research questions:

- **Can a smartphone be used as a plethysmograph in order to accurately determine heart rate and blood pressure?**

  Based on the results achieved in this dissertation, this appears to be a promising solution. The results for the BP algorithm may not yet be up to the standards for commercial sphygmomanometers, but are close enough to provide a reasonable indication that could soon reach close to the accuracy of cuff-based methods.

- **Is it necessary to determine all the characteristic risk factors for MetS in order to diagnose metabolic syndrome?**

  The MetS algorithm did not achieve the accuracy that may be required of a diagnostic tool, however, it is capable of providing a risk indicator, which could potentially be used to determine whether further diagnostics are required. In that sense, it falls perfectly in line with what would be expected from a screening solution.

- **Can information from lifestyle factors, such as diet, exercise or smoking, be used to improve the accuracy of MetS predictions?**

There was 2.9% accuracy increase seen in the MetS model when including lifestyle factors and medical history, which indicates that there is value in including lifestyle habits to screen for MetS that may not be included in the standard definitions.

- **Would an application of this kind be easy enough to use for it to be a feasible solution?**

  Based on all three usability metrics, the application performed well, showing that an application of this kind could be usable enough to be a feasible screening solution.

This means that when answering "How accurately could a smartphone screen for metabolic syndrome, and would this be a feasible solution?", one could answer that based on the results from this dissertation an application of this kind could screen for MetS with an accuracy 73%, which could be further improved depending on the use environment and confidence levels required. Based on these results, we can say that smartphone based metabolic syndrome screening is a promising venture, that can be expanded upon in further research and considered for real world screening solutions.

BIBLIOGRAPHY

[1] J. Kaur, "A Comprehensive Review on Metabolic Syndrome," *Cardiology Research and Practice*, vol. 2014, no. 943162, pp. 1–24, 2014.

[2] S. Grundy *et al.*, "Diagnosis and Management of the Metabolic Syndrome: An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement," *American Heart Association*, vol. 10, no. 1161, pp. 2735–2752, 2005. DOI: 10.1161/CIRCULATIONAHA.105.169404.

[3] P. Pérez-Martínez *et al.*, "Lifestyle recommendations for the prevention and management of metabolic syndrome: an international panel recommendation," *Nutrition Reviews*, vol. 75, no. 5, pp. 307–326, 2017.

[4] J. Kaur, "Assessment and Screening of the Risk Factors in Metabolic Syndrome," *Medical Sciences*, vol. 2, pp. 140–152, 2014. DOI: 10.3390/medsci2030140.

[5] H. Yanai, "Metabolic Syndrome and COVID-19," *Cardiology Research*, vol. 11, no. 6, pp. 360–365, 2020, ISSN: 1923-2829. DOI: 10.14740/cr1181. [Online]. Available: http://www.cardiologyres.org/index.php/Cardiologyres/article/view/1181.

[6] A. Worachartcheewan, V. Prachayasittikul, and C. Nantasenamat, "Data mining for the identification of metabolic syndrome status," *EXCLI Journal*, vol. 2018, no. 17, pp. 72–88, 2018.

[7] A. Wagh and N. Stone, "Treatment of metabolic syndrome," *Expert Review Cardiovascular Therapy*, vol. 2, no. 2, pp. 213–228, 2004.

[8] R. Erasmus *et al.*, "High prevalence of diabetes mellitus and metabolic syndrome in a South African coloured population: Baseline data of a study in Bellville, Cape Town," *The South African Medical Journal*, vol. 102, no. 11, pp. 841–844, 2012.

[9] Centers for Disease Control and Prevention, *Top 10 Causes of Death*, https://www.cdc.gov/globalhealth/countries/southafrica/default.htm, 2018.

[10] N. Barakat, A. Bradley, and M. Barakat, "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 1114–1120, 2010.

[11] N. Barakat, "Diagnosis of Metabolic Syndrome: A Diversity Based Hybrid Model," in *Machine Learning and Data Mining in Pattern Recognition*, 2016, pp. 185–198.

[12] Y. Ushida *et al.*, "Combinational risk factors of metabolic syndrome identified by fuzzy neural network analysis of health-check data," *BMC Medical Informatics and Decision Making*, vol. 12, no. 80, 2012.

[13] H. Park and S. Cho, "Evolutionary attribute ordering in Bayesian networks for predicting the metabolic syndrome," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4240–4249, 2012. DOI: 10.1016/j.eswa.2011.09.110.

[14] A. Worachartcheewan and C. Nantasenamat, "Identification of metabolic syndrome using decision tree analysis," *Diabetes Research and Clinical Practice*, vol. 90, no. 1, e15–e18, 2010, ISSN: 0168-8227. DOI: 10.1016/j.diabres.2010.06.009.

[15] D. Ivanović, A. Kupusinac, E. Stokić, R. Doroslovački, and D. Ivetić, "ANN Prediction of Metabolic Syndrome: a Complex Puzzle that will be Completed," *Journal of Medical Systems*, vol. 40, no. 264, 2016.

[16] Y. Kim and Y. Je, "Meat Consumption and Risk of Metabolic Syndrome : Results from the Korean Population and a Meta-Analysis of Observational Studies," *Nutrients*, vol. 10, no. 390, pp. 1–16, 2018. DOI: 10.3390/nu10040390.

[17] C. Pitsavos, D. Panagiotakos, M. Weinem, and C. Stefanadis, "Diet, Exercise and the Metabolic Syndrome," *Journal of the Society of Biomedical Diabetes Research*, vol. 3, no. 3, pp. 118–126, 2006.

[18] Myfitnesspal, *Free Calorie Counter, Diet and Exercise Journal*, https://www.myfitnesspal.com/.

[19] Cronometer, *Cronometer: Track Nutrition and Count Calories*, https://cronometer.com/.

[20] M. Sadeghi, M. Gharipour, P. Nezafati, D. Shafie, E. Aghababaei, and N. Sarrafzadegan, "Assessing Metabolic Syndrome Through Increased Heart Rate During Exercise," *Acta Med Iran*, vol. 54, no. 11, pp. 724–730, 2016.

[21] P. Yan *et al.*, "Resting and Postexercise Heart Rate Detection From Fingertip and Facial Photoplethysmography Using a Smartphone Camera: A Validation Study," *JMIR Mhealth Uhealth*, vol. 5, no. 3, e33, Mar. 2017. DOI: 10.2196/mhealth.7275.

[22] O. Rogowski *et al.*, "Elevated resting heart rate is associated with the metabolic syndrome," *Cardiovascular Diabetology*, vol. 8, no. 55, 2009.

[23] J. S Rana *et al.*, "Resting Heart Rate and Metabolic Syndrome in Patients With Diabetes and Coronary Artery Disease in Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) Trial," *Preventive Cardiology*, vol. 10, no. 111, pp. 112–116, 2010. DOI: 10.1111/j.1751-7141.2010.00067.x.

[24] C. Phillips *et al.*, "Obesity and body fat classification in the metabolic syndrome: impact on cardiometabolic risk metabotype.," *The Obesity Society (Silver Spring)*, vol. 21, no. 1, 2013.

[25] M. Elgendi, "On the Analysis of Fingertip Photoplethysmogram Signals," *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[26] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "Smartphone-Based Photoplethysmogram Measurement," in *Digital Image, Signal and Data Processing for Measurement Systems*, River Publishers, 2012, pp. 135–164.

[27] X. Xing and M. Sun, "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks," *Biomedical Optics Express*, vol. 7, no. 8, pp. 3007–3020, 2016.

[28] I. Jeong, S. Jun, D. Um, J. Oh, and H. Yoon, "Non-Invasive Estimation of Systolic Blood Pressure and Diastolic Blood Pressure Using Photoplethysmograph Components," *Yonsei Medical Journal*, vol. 51, no. 3, pp. 345–353, 2010.

[29] S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani, "Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method," *Biomedical Signal Processing and Control*, vol. 47, pp. 196–206, 2019, ISSN: 17468108. DOI: 10.1016/j.bspc.2018.08.022.

[30] K. Matsumura, P. Rolfe, S. Toda, and T. Yamakoshi, "Cuffless blood pressure estimation using only a smartphone," *Scientific Reports*, vol. 8, no. 7298, pp. 1–10, 2018, ISSN: 20452322. DOI: 10.1038/s41598-018-25681-5.

[31] C. Herman and H. Gill, "Screening for Preclinical Disease: Test and Disease Characteristics," *Fundamentals of Clinical Research for Radiologists*, vol. 179, no. 4, pp. 825–831, 2002.

[32] World Health Organization, "mHealth: New horizons for health through mobile technologies," *Global Observatory for eHealth series*, vol. 3, 2011.

[33] L. D. Maxim, R. Niebo, and M. J. Utell, "Screening tests: a review with examples," *Inhalation Toxicology*, vol. 26, no. 13, pp. 811–828, 2014, ISSN: 0895-8378. DOI: 10.3109/08958378.2014.955932. [Online]. Available: https://www.tandfonline.com/doi/full/10.3109/08958378.2014.955932.

[34] A Morabia, "History of medical screening: from concepts to action," *Postgraduate Medical Journal*, vol. 80, no. 946, pp. 463–469, 2004, ISSN: 0032-5473. DOI: 10.1136/pgmj.2003.018226. [Online]. Available: https://pmj.bmj.com/lookup/doi/10.1136/pgmj.2003.018226.

[35] B. De Ridder, B. Van Rompaey, J. Kampen, S. Haine, and T. Dilles, "Smartphone Apps Using Photoplethysmography for Heart Rate Monitoring: Meta-Analysis," *JMIR Cardio*, vol. 2, no. 1, e4, 2018, ISSN: 2561-1011. DOI: 10.2196/cardio.8802.

[36] M. Poh *et al.*, "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart BMJ Journals*, Epub ahead of print, 2018.

[37] P. Chan *et al.*, "Performance of a Smartphone-Based Photoplethysmographic Application for Atrial Fibrillation Screening in a Primary Care Setting," *Journal of the Medical Heart Association*, vol. 5, no. 7, e003428, 2016.

[38] A. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.

[39] A. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 160035, 2016.

[40] T. Lin, Y. Mayzel, and K. Bahartan, "The accuracy of a non-invasive glucose monitoring device does not depend on clinical characteristics of people with type 2 diabetes mellitus," *Journal of Drug Assessment*, vol. 7, no. 1, pp. 1–7, 2018.

[41] DiaMonTech GmbH, *DiaMonTech: non-invasive glucose measurements*, `http://www.diamontech.de/`, 2017.

[42] Medical Wireless Sensing Ltd, *GlucoWise : Meet the new non-invasive glucose monitor that helps you take control of your life*, `http://www.gluco-wise.com/`, 2018.

[43] T. Minh *et al.*, "Noninvasive Measurement of Plasma Triglycerides and Free Fatty Acids from Exhaled Breath," *Journal of Diabetes Science and Technology*, vol. 6, no. 1, pp. 86–101, 2012.

[44] S. Shimawaki, Y. Kobayashi, M. Nakabayashi, and N. Sakai, "Non-invasive Serum Cholesterol Detection Using Near-infrared Light Transmission," *Biomedical Engineering Research*, vol. 3, no. 3, pp. 80–87, 2014.

[45] E. Aristovich, *Non-invasive Measurement of Cholesterol in Human Blood by Impedance Technique: an Investigation by Finite Element Field Modelling*. 2014, pp. 20–23.

[46] R. Schutte *et al.*, "Blood glutathione and subclinical atherosclerosis in African men: The SABPA study," *American Journal of Hypertension*, vol. 22, no. 11, pp. 1154–1159, 2009.

[47] S. Jahandideh, M. Jahandideh, S. Asefzadeh, and A. Ziaee, "The Use of Artificial Neural Network and Logistic Regression to Predict the Influence of Lifestyle on Cardiovascular Risk Factors," *COJ Nurse Healthcare*, vol. 1, no. 1, pp. 1–5, 2017. DOI: `10.31031/COJNH.2017.01.000505`.

[48] ISO 9241-11:2018(en), "Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts," International Organization for Standardization, Tech. Rep. 11, 2018.

[49] J. Sauro and E. Kindlund, "A Method to Standardize Usability Metrics Into a Single Score," in *SIGCHI Conference on Human Factors in Computing Systems*, 2005, pp. 401–409.

[50] J. Brooke, "SUS : A Retrospective," *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, 2013.

[51] D. Pham and X. Liu, *Neural Networks for Identification, Prediction and Control*. 2012, pp. 1–21.

[52] G. Sanderson, *Neural networks*, 3Blue1Brown. `https://youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi`, 2018.

[53] J. Allen, "Photoplethysmography and its application in clinical physiological measurement.," *Physiological Measurement*, vol. 28, no. 3, 2007.

[54] F. Tabei, J. M. Gresham, B. Askarian, K. Jung, and J. W. Chong, "Cuff-less blood pressure monitoring system using smartphones," *IEEE Access*, vol. 8, pp. 11 534–11 545, 2020, ISSN: 2169-3536. DOI: `10.1109/ACCESS.2020.2965082`.

[55] J. Liu, S. Qiu, N. Luo, *et al.*, "PCA-Based Multi-Wavelength Photoplethysmography Algorithm for Cuffless Blood Pressure Measurement on Elderly Subjects," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 663–673, 3 Mar. 2021, ISSN: 2168-2194. DOI: 10.1109/JBHI.2020.3004032.

[56] Anaesthesia UK, "Principles of pulse oximetry," Tech. Rep., 2004.

[57] D. Grimaldi, Y. Kurylyak, F. Lamonaca, and A. Nastro, "Photoplethysmography detection by smartphone's videocamera," in *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, vol. 1, 2011, pp. 488–491. DOI: 10.1109/IDAACS.2011.6072801.

[58] E. Jonathan and M. Leahy, "Investigating a smartphone imaging unit for photoplethysmography," *Physiological Measurement*, vol. 31, no. 11, pp. 79–83, 2010.

[59] E. Jonathan and M. Leahy, "Cellular phone-based photoplethysmographic imaging," *Journal on Biophotonics*, vol. 4, no. 5, pp. 293–296, 2011.

[60] J. E. Hall and A. C. Guyton, *Guyton and Hall Textbook of Medical Physiology*, 13th ed. 2015, p. 114, ISBN: 978-1-4557-7005-2.

[61] K. Naik and P. H. Bhathawala, "Mathematical Modeling of Human Cardiovascular System: A Lumped Parameter Approach and Simulation," *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, vol. 11, no. 2, pp. 73–84, 2018.

[62] B. Moody, G. Moody, M. Villarroel, G. Clifford, and I. Silva, *MIMIC-III Waveform Database (Version 1.0)*, 2020. DOI: https://doi.org/10.13026/c2607m.

[63] Kent State University, *Pearson Correlation*, https://libguides.library.kent.edu/SPSS/PearsonCorr, 2021.

[64] S. Ruder, "An overview of gradient descent optimization algorithms," Insight Centre for Data Analytics, Dublin, Tech. Rep., 2017, pp. 1–14.

[65] C. Lin, Y. Bai, J. Chen, T. Hwang, C. T.T., and H. Chiu, "Easy and low-cost identification of metabolic syndrome in patients treated with second-generation antipsychotics: artificial neural network and logistic regression models.," *J Clin Psychiatry*, vol. 71, no. 3, pp. 225–34, 2010.

[66] M. Mifflin, S. Jeor, L. Hill, B. Scott, S. Daugherty, and Y. Koh, "A new predictive equation for resting energy expenditure in healthy individuals," *The American Journal of Clinical Nutrition*, vol. 51, no. 2, pp. 241–247, 1990.

[67] World Health Organization, "WHO Technical specifications for automated non-invasive blood pressure measuring devices with a cuff," Tech. Rep., 2020.

[68] G. Stergiou *et al.*, "Recommendations and Practical Guidance for performing and reporting validation studies according to the Universal Standard for the validation of blood pressure measuring devices," Tech. Rep., 2019.

[69] Department of Health RSA, *Ethics in Health Research: Principles, Processes and Structures*, 2015.

[70] Omron Healthcare, *HEM-712C Automatic Blood Pressure Monitor with Arm Cuff*, 2005.

[71] B. Nelson and N. Allen, "Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study," *JMIR mHealth and uHealth*, vol. 7, no. 3, e10828, 2019.

[72] M. Nelson, L. Kaminsky, D. Dickin, and A. Montoye, "Validity of consumer-based physical activity monitors for specific activity types," *Med Sci Sports Exerc*, vol. 48, no. 8, pp. 1619–28, 2016.

[73] S. Stahl, H. An, D. Dinkel, J. Noble, and J. Lee, "How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough?" *BMJ Open Sport Exerc Med.*, vol. 2, no. 1, e000106, 2016.

[74] Consumer Technology Association, *Physical Activity Monitoring for Heart Rate ANSI/CTA-2065*, `https://standards.cta.tech/kwspub/published_docs/CTA-2065-Preview.pdf`, 2018.

[75] US Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER) Food and Drug Administration, *Guidance for Industry, Investigating Out-of-Specification (OOS), Test Results for Pharmaceutical Production*, `https://www.fda.gov/downloads/drugs/guidances/ucm070287.pdf`, 2006.

[76] A. Shcherbina, C. Mattsson, D. Waggott, *et al.*, "Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort," *Journal of Personalized Medicine*, vol. 7, no. 2, p. 3, 2017, ISSN: 2075-4426. DOI: `10.3390/jpm7020003`. [Online]. Available: `http://www.mdpi.com/2075-4426/7/2/3`.

[77] Python Software Foundation, *Generate pseudo-random numbers*, The Python Standard Library. `https://docs.python.org/3.6/library/random.html`, 2021.

[78] USBR, "Procedure for Calibrating Balances or Scales," pp. 38–41, 1989. [Online]. Available: `https://www.usbr.gov/tsc/techreferences/mands/rockmanual/tan_earthmanualUSBR/USBR1012.pdf`.

APPENDIX

A

METABOLIC SYNDROME FEATURES

This appendix contains a breakdown of the features from the SABPA dataset that were used to train and assess the MetS model. Tables A.1 and A.2 give a description of each variable that was used and the method in which it was measured, where applicable.

Table A.1: Description of variables 1 to 21 from SABPA dataset applicable to metabolic syndrome

Variable	Description	Measurement method
MetS	Metabolic Syndrome	Modified NCEP ATPIII method
Age	Participant age	
Race	Participant race / ethnicity	
Sex	Male / Female	
CVD History	Cardiovascular disease history (Disease affecting your heart or blood vessels)	
Stroke	History or incidence of stroke	
MI/cardiac events	Myocardial infarction (i.e. heart attack) or cardiac events history	
Kidney disease	History of kidney disease	
Atrial fibrillation	History of atrial fibrillation	
BP drugs	Use of anti-hypertensive drugs	
Anti-diabetic drugs	Use of anti-diabetic drugs	
Stature	Participant's stretched stature in cm to the nearest 0.1cm	Measured with a stadiometer while the participants head is in the Frankfort plane (standard anatomical position)
Body Mass	Body Mass in kg measured to the nearest 0.1 kg while wearing minimal clothes	Electronic digital scale. Calibration according to USBR[1] 1012-89 [78]
BMI	BMI calculated as kg/m^2	
Waist	Waist circumference in cm	Measured at the midpoint between the lower costal border and the iliac crest perpendicular to the long axis of the trunk
TEE	Total Energy Expenditure (kcal) in 24 h, taking Resting Metabolic Rate (RMR) into account	Actical® activity monitor
Smoking	Current smoking habit	Selfreporting questionnaire
Alcohol	Current consumption of alcoholic beverages	Selfreporting questionnaire
Glucose	Fasting venous blood glucose in mmol/l	One Touch Sure Step™ Glucose meter
HdL chol	Serum high density lipoprotein cholosterol in mmol/l.	Timed-end-point method, Unicel DXC 800
TRIG	Serum triglycerides in mmol/l.	Timed-end-point method, Unicel DXC 800

Table A.2: Description of variables 22 to 24 from SABPA dataset applicable to metabolic syndrome

Variable	**Description**	**Measurement method**
SBP	Resting sphygmomanometer systolic blood pressure	Littmann® II S.E. Stethoscope 2205 and Riester CE 0124 Manual Sphygmomanometer
DBP	Resting sphygmomanometer diastolic blood pressure	Littmann® II S.E. Stethoscope 2205 and Riester CE 0124 Manual Sphygmomanometer
Rest HR	Mean resting heart rate, beats per minute	Finometer® ECG module