A GRADIENT BOOSTING REGRESSION MODEL FOR THE PREDICTION OF AN INDIVIDUAL'S SHORT TERM BLOOD PRESSURE

KIPTOO DAISY JEPKOECH

A Thesis Submitted to the Institute of Postgraduate Studies of Kabarak University in Partial Fulfillment of the Requirements for the Award of Master of Science in Information Technology

KABARAK UNIVERSITY

NOVEMBER, 2023

DECLARATION

1. I do hereby declare that:

   i. This thesis is my own work and to the best of my knowledge, it has not been presented for the award of a degree in any university or college.

   ii. That the work has no incorporated material from other works or a paraphrase of such material without due and appropriate acknowledgement

   iii. That the work has been subjected to processes of anti-plagiarism and has met Kabarak University 15% similarity index threshold

2. I do understand that issues of academic integrity are paramount and therefore I may be suspended or expelled from the University or my degree may be recalled for academic dishonesty or any other related academic malpractices

Signature ……………………………………… Date ………………………….

Kiptoo Daisy Jepkoech

GMI/NE/0386/05/17

RECOMMENDATION

To the Institute of Postgraduate Studies:

The thesis entitled "**A Gradient Boosting Regression Model for the Prediction of an Individual's Short Term Blood Pressure**" and written by **Kiptoo Daisy Jepkoech,** is presented to the Institute of Postgraduate Studies of Kabarak University. We have reviewed the thesis and recommend it be accepted in partial fulfillment of the requirement for award of the Master of Science in Information Technology.

Signature………………………………… Date……………………………

Dr. Moses Thiga

Senior Lecturer, School of Science, Engineering and Technology,

Kabarak University

Signature………………………………… Date……………………………

Dr. Peter Rugiri

Senior Lecturer, School of Science, Engineering and Technology,

Kabarak University

Signature………………………………… Date……………………………

Dr. Pamela Kimeto

Senior Lecturer, School of Medicine and Health Sciences,

Kabarak University

COPYRIGHT

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisors Dr. Moses Thiga, Dr. Peter Rugiri and Dr. Pamela Kimeto for their professional advice and help throughout this thesis. Also, my sincere thanks goes to my classmates and friends for their support throughout the study. I sincerely thank my parents for their continued support both financially and spiritually which made this study a success. Finally, I must express heartfelt gratitude to the Almighty God for the wisdom and wellbeing.

DEDICATION

I dedicate this study to my parents; Prof. Ezekiel Kiprop and the late Gladys Birgen, my siblings; Peter Kiptoo, Phyllis Kiptoo, Millicent Kiptoo, Ivy Kiptoo and Lexy Kiptoo.

ABSTRACT

Hypertension is a serious problem across the globe because of its high mortality rate per year. High Blood Pressure (BP) has no warning signs nor symptoms, and measuring the BP level is the only way to know about a person's BP status. Many solutions geared toward managing hypertension have been successful, but the condition still persists across the globe. Though there is treatment to help those with hypertension manage the condition, there is a lack of a suitable solution to predict a person's BP based on previous readings and planned future activities. This study took a different approach to this problem through the use of Artificial Intelligence (AI), Machine Learning (ML) in particular. An ML model was used to predict future BP fluctuations of an individual's BP using their future calendar events. The study was done in Uasin-Gishu County, Kenya.The researcher employed design science and experimental methods for the study. Rapid Application Development was used in order to design the smartphone application that captured the data from the individuals. The data for the study was collected using a smartwatch, which collected the BP and heartrate and a smartphone application which collected the mood, activities and calendar events of the individuals. The Gradient Boosting Regression predictive model was implemented using the Iterative and Incremental Development Model. The Holdout method's, test dataset was used along with R-Squared (R^2) and Mean Absolute Error (MAE) to evaluate the prototype. The ML model gave an accuracy score of 99% in predicting an individual's BP. The study also reviled some relationships among the attributes that were used, an example is the relationship between the individuals' activities and their BP. From the findings of the study, it is recommended that further studies apply these findings to create custom informative notifications through a phone application, email or Short Message Service (SMS) for each individual in order to prevent hypertension or even lower BP in case of a hypertensive patient.

Keywords: *Blood Pressure, Hypertension, Artificial Intelligence, Machine Learning, Gradient Boosting Regression, Prediction*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AI	Artificial Intelligence
ANNs	Artificial Neural Networks
BLE	Bluetooth Low Energy
BP	Blood Pressure
GHO	Global Health Observatory
GPS	Global Positioning System
IDE	Integrated Development Environment
IT	Information Technology
KEMRI	Kenya Medical Research Institute
ML	Machine Learning
MTRH	Moi Teaching and Referral Hospital
PPG	Photo Plethysmography
SBP	Systolic Blood Pressure
DBP	Diastolic Blood Pressure
SDG	Sustainable Development Goals
WHO	World Health Organization
GBR	Gradient Boosting Regressor
CHD	Coronary Heart Disease
BMI	Body Mass Index
HR	Heart Rate
DM	Diabetes Mellitus
HLP	Hyperlipidemia
CKD	Chronic Kidney Disease
EMD	Emotional or Mental Disorders
SlpD	Sleep Disorders
WC	Waist Circumference
HC	Hip Circumference
CART	Classification and Regression Tree
CPCSSN	Canadian Primary Care Sentinel Surveillance Network

OPERATIONAL DEFINITION OF TERMS

Artificial Intelligence The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings (Copeland, 2019).

Machine Learning This enables machines to learn without programming them explicitly (Mohammed *et al.,* 2017).

Artificial Neural Networks They are brain-inspired systems which are intended to replicate the way that we humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use (Dormehl, 2019).

Bluetooth Low Energy Bluetooth Low Energy is the intelligent, power-friendly version of Bluetooth wireless technology (Gupta &Mohammed, 2016).

Prediction This refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when you're trying to forecast the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days (Prediction, n.d).

Integrated Development Environment It provides interfaces for users to write code, organize text groups, and automate programming redundancies (Walker, 2018).

Java This is a high-level computer programming language. It enables programmers to write instructions using English-based commands instead of having to write in numeric codes.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

This chapter presents the background to the study that is blood pressure, hypertension and its risks; complications, and its management, statement of the problem, the purpose of the study, and objectives of the study further it presents the research questions, the justification, scope and assumptions of the study.

1.2 Background of the Study

Blood pressure (BP) is the force of blood pushing against the walls of a person's arteries, as it carries blood from the heart to other parts of the body. BP normally rises and falls throughout the day, if it stays high for a long time, it can damage the heart and lead to health problems (Merai, Siegel, Rakotz, Basch, Wright, Wong & Thorpe, 2016).

1.2.1 Hypertension

Table 1

Classification of blood pressure based on systolic and diastolic readings

Blood Pressure Classification	Systolic Blood Pressure (SBP)	Diastolic Blood pressure (DBP)
Normal	<120	And <80
Prehypertension	120-139	or 80-89
Stage 1 hypertension	140-159	or 90-99
Stage 2 hypertension	≥ 160	≥ 100

Note: Tran and Giang (2013)

Table 1 shows how BP can be classified according to Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP) of an individual, in to Normal, Pre-hypertension, Stage 1 hypertension and Stage 2 hypertension. The SBP indicates how much pressure is in the arteries when the heart beats. The DBP measures the pressure in the arteries in between heartbeats.

Hypertension is another term for High Blood Pressure. Hypertension has no warning signs nor symptoms, and many people do not know they have it. Measuring the BP level is the only way to know about a person's BP status. One can then take steps to control it if it is high. High BP can seriously hurt important organs like the kidneys, heart and brain. In addition to coronary heart diseases and stroke, complications of raised BP include heart failure, peripheral vascular disease, renal impairment, retinal hemorrhage and visual impairment. It is also a major cause of clinical and pre-clinical damage to the brain, kidneys, and arterial blood vessels (Merai *et al.*, 2016). According to the World Health Organization (WHO), ischemic heart disease (also referred to as coronary heart disease) and stroke are the world's biggest killers, accounting for a combined 15.2 million deaths in 2016 (Nichols, 2017). The two conditions can be derived from hypertension.

1.2.2 Hypertension Management

Hypertension is managed using lifestyle modification, antihypertensive medications, reducing and managing mental stress and also the regular measurement of BP (Kitt, Fox, Tucker & McManus, 2019). Hypertension is usually treated to achieve a BP of less than or equal to 130/90 mmHg to 160/100 mmHg. Lifestyle change is usually recommended as the first line of treatment. It includes; dietary changes like reducing salt intake and eating more fruit and vegetables, physical exercise and weight loss. If the patient’s hypertension is too high, medication is usually recommended alongside the lifestyle changes (Al-Riyami & Nadar, 2021).

A key management activity for BP is its regular measurement and current approaches to measure BP include: Automated office BP measurement which is the average of multiple

BP readings recorded with a fully automated device with the patient resting silently, alone, in a quiet room of an office or clinic. For hypertension to be diagnosed, the patient has to be measured more than three times, on different occasions. The disadvantage of relying only on measuring BP in a clinical setting is that people cannot find time to go to a clinic and measure their BP levels on a daily basis. This can sometimes lead to wrong diagnosis and therefore incorrect treatment. There is also the problem of the white-coat effect, (which means that BP is higher when it is taken in a medical setting than it is when taken at home) (Whelton et al., 2018).

24-hour ambulatory blood pressure monitoring is another approach for monitoring BP that measures BP at regular intervals. It has a greater number of readings, the absence of observer bias, and the reduction of white-coat effect. There is also self-monitoring BP at home or on the go by using BP monitors, which also has a greater number of readings and minimizes the white-coat effect, contributing to better diagnostic accuracy, compared with conventional sphygmomanometer (Staessen *et al.,* 2017). This method as research has shown, Staessen-Li *et al*., (2017) produces better results. Its disadvantage is that the BP monitors may be expensive and cumbersome, also, one may forget to measure their BP when needed.

1.2.3 Devices Used to Measure Blood Pressure and the Monitoring of Blood Pressure

There are several BP measurement devices. They include; manual sphygmomanometers, digital non-portable for upper arm with automatic inflation, digital portable for upper arm with automatic inflation, digital portable for wrist with automatic inflation, and digital portable for finger with automatic inflation, and blood pressure watches (Kracher, 2019). Sphygmomanometer and stethoscope are traditional office BP measurement devices. These devices are being phased out and replaced by other oscillometric devices because of bulkiness and mercury's toxicity.

Wrist and finger devices are small and portable which is a big advantage. This means that BP can be measured in almost all situations, even during exercises like swimming if the device is water proof. Some of these devices, especially the smartwatches, have their own mobile applications. These applications can be used to measure BP, monitor the dynamics of BP changes using graphs and statistics, share the information with a doctor, and also give reminders to the patients in order for them to take their medication. The measurements taken via these applications can be used to serve a greater purpose by using the data collected for purposes of BP predictions.

1.2.4 Prediction of BP as a Management Strategy

One way of improving the management of hypertension is through predicting the probability of a person becoming hypertensive or the probability of the BP rising above or falling below normal levels (Kitt *et al.*, 2019). Using data collected from individuals together with machine learning, some studies have focused on predicting the onset of hypertension as a management strategy. The studies show that prediction has been done mainly to diagnose if the individual is, at that current moment, hypertensive or not, or determine if he or she is prone to have hypertension in the future. Three examples have been provided:

Ye *et al.*, (2018) developed a risk prediction model of incident essential hypertension within the next year. Data was collected from individual patient electronic health records. XGBoost machine learning algorithm was employed in the process of feature selection and model building. The one-year risk prediction model was accurate, it attained Areas Under the Curve (AUCs) of 0.917 and 0.870 in the retrospective and prospective cohorts, respectively. Ye *et al.*, (2018) established that the model would help in hypertension care in that, some of the high-risk individuals would receive diagnosis in the first 3 months of the next year therefore that would enable monitoring so as to prevent or delay incident hypertension. A study done by LaFreniere *et al.,* (2016), identified the important risk factors based on patients' health conditions, medical records, and demographics which were then used to predict the presence of hypertension in an individual using Artificial Neural Networks (ANNs). The model could also predict the probability of a person developing hypertension in the future with about 82% accuracy and can, therefore, be used as an early warning system. In 2018, a study was done by Patnaik *et al.,* (2018) to predict the occurrence of essential hypertension using annual health records. The main objective was to predict if an individual will have hypertension in the next year.

1.2.5 BP Prediction Techniques

A variety of techniques have been used in the prediction of BP. These include prediction of hypertension using machine learning, association rule learning and artificial intelligence. A study done by Zhang and colleagues in 2019 used machine learning to predict hypertension. Support Vector Regression (SVR) algorithm was used to investigate the implicit association between human physiological index data and BP measurements collected by medical devices to obtain an efficient and accurate prediction model for BP (Zhang *et al.,* 2019). Another research by Liu *et al.,* in 2018 did a study on Identification of Hypertension by Mining Class Association Rules from Multi-Dimensional Features in 2018. Their objective was a class association rules-based method to identify hypertension. This was to utilize the relationship existing in multi-dimensional features to characterize hypertension pattern more effectively, in order to improve the identification performance. Using two neural network algorithms, back-propagation and radial basis function, an experiment done by Kwong *et al.,* (2016) found that the Artificial Neural Networks (ANNs) were able to predict SBP with an accuracy of over 90%.

Artificial Neural Networks, Classification Models, Regression Models, Association Rule Learning and Adaptive neuro fuzzy inference systems to name a few, are some of the various approaches that aid in the prediction of hypertension. Most studies show how machine learning has been used in the diagnosis of the condition or to predict if a person will be susceptible to high BP in the future. "To date, AI has been mainly used to investigate risk factors for hypertension but has not yet been utilized for hypertension management due to the limitations of study design and of physician's engagement in computer science" (Krittanawong *et al.*, 2018). Krittanawong *et al.*, (2018) further state that wearable technology will play a big role in enhancing AI abilities to be able to predict and therefore help in the control of an individual's blood pressure.

Hypertension is a silent disease that is only diagnosed following a series of blood pressure measurements thus the need for continuous monitoring and prediction for those at risk.

1.3 Statement of the Problem

BP varies throughout the day and can change with the physical and psychological state of an individual, their environment, the observer, and the measuring device. If an individual’s BP level begins to rise consistently, and action is not taken due to factors such as lack of awareness, the person might develop high BP. Complications in the body might occur if high BP is not dealt with; an individual may develop problems with the brain, retina and arteries to name a few. Due to the pervasiveness of hypertension, individualized prediction of BP is very important as it may significantly reduce the rise of BP by giving precautions to an individual. The study helped in predicting an individual’s short term BP based on their BP readings and activities. It also helped to pin-point why an individual’s BP was raised. This study uses a smart watch to collect BP, a smartphone application to collect activity data that is needed inclusive of the BP, and a machine learning model that predicts future fluctuations of an individual’s BP using the data provided.

1.4 Purpose of the Study

The purpose of this study was to develop a machine learning solution that could predict an individual’s future BP level based on their past BP readings and activity data.

1.5 The Specific Objectives of the Study

i. To develop a model for regular collection of blood pressure readings and activity data.

ii. To investigate a suitable machine learning algorithm for the prediction of future blood pressure readings of an individual using their past blood pressure readings and activity data.

iii. To train the suitable machine learning algorithm to predict future blood pressure readings of an individual using their past blood pressure readings and activity data.

iv. To validate the performance of the machine learning algorithm for learning and predicting future levels of an individual's blood pressure using their past readings and activity data.

1.6 Research Questions

i. What is a suitable model for the development of a system for regular collection of blood pressure readings and activity data?
ii. Which machine learning algorithm will be suitable for the prediction of future blood pressure readings of an individual using their past blood pressure readings and activity data?
iii. How will the suitable machine learning algorithm be trained to predict future blood pressure readings of an individual using their past blood pressure readings and activity data?
iv. How the machine learning algorithm does for learning and predicting future levels of an individual's blood pressure using their past readings and activity data perform?

1.7 Justification for the Study

Raised BP is a prevalent health condition often characterized by a lack of awareness among a substantial proportion of affected individuals. This study effectively enhanced individuals' awareness of their BP status, thus holding significant potential for substantial improvements in BP monitoring and control through predictive mechanisms. Furthermore, the study shed light on the specific activities that contribute to elevated BP levels on an individual basis. These findings not only contribute to individual well-being but also provide a valuable evidence-based approach that enhances policymaking efficacy and leads to more favorable policy outcomes.

1.8 Scope of the Study

The study focused on these main objectives: Firstly, it aimed to gather both blood pressure (BP) and activity data. Secondly, it aimed to utilize this collected data to generate individual BP predictions through the application of machine learning techniques. However, it is important to note that the study did not proceed to the complete implementation of the trained AI model onto a mobile application. Providing feedback to individuals was not included in the scope of work for this particular study. The process of deploying an AI model onto a mobile platform is known to be time-consuming, and due to the limited duration of the study, this aspect could not be accommodated. The graphical scope of the study was Uasin Gishu County and all its constituencies.

1.9 Assumptions of the Study

The assumptions of the study were that;

i. The participants have a smartphone with Bluetooth capability.
ii. The participants are conversant with the operation of a calendar application.
iii. The calendar events that the participants will add to the calendar application will be consistent.
iv. The participants are willing to wear the smartwatch required for BP data collection.

The following mitigation measures were considered;

i. Smartphone with Bluetooth Capability: For the participants who were willing to participate in the study and had no smartphones, they borrowed one from their friends or family.
ii. Familiarity with Calendar Application: Training sessions on how to use the calendar application ensured participants were proficient in its operation. Technical support and assistance was also offered throughout the study to address any difficulties participants encountered.

iii. Consistency in Calendar Events: Clearly defined and communicated specific guidelines for adding calendar events were provided to ensure consistency among participants.

iv. Willingness to Wear Smartwatch: The benefits and importance of wearing the smartwatch for accurate blood pressure (BP) data collection were emphasized to encourage participant compliance.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter discusses hypertension, an overview of smartwatches and their use in the measurement of BP, BP applications, the use of machine learning to predict BP, and the theoretical and conceptual frameworks.

2.2 Hypertension

According to the Global Health Observatory (GHO) data from the World Health Organization (WHO), one of the leading risks factors for global mortality is raised BP. This is due to the complications that hypertension brings about such as, heart failure, peripheral vascular disease, renal impairment, retinal haemorrhage, visual impairment, coronary heart diseases and stroke (Shackelford, 2019). Raised BP affects 1.13 billion people worldwide and it is predicted to be increased to 1.56 billion adults by the year 2025 (Singh *et al.*, 2017).

The United Nations Development Programs, Sustainable Development Goal (SDG) number three, is good health and well-being. According to the (SDG), health is a driver, indicator and outcome of sustainable development and that healthy people are better able to contribute to the development of their countries. According to their facts and figures every 2 seconds, someone aged 30 to 70 years dies prematurely from non-communicable diseases, these are; cardiovascular disease, chronic respiratory disease, diabetes or cancer (Mohammed & Ghebreyesus, 2018). Hypertension, thus, has to be significantly managed, being part of the non-communicable diseases causing cardiovascular diseases.

One of the objectives of the Medical Research arm of the Government, Kenya Medical Research Institute (KEMRI), is “to cooperate with other research organizations and institutions of higher learning on matters of relevant research and training" (Agata, 2018). One of their six main research programs that are aligned to the KEMRI Strategic Master Plan and the Vision 2030 is to conduct basic, clinical, operational, implementation and applied research in all matters related areas to non-communicable diseases such as cancer, diabetes, lifestyle diseases and mental health. The flagship project areas that the institute wants to address are: Life Style Diseases, Obesity, Diabetes, Hypertension, Drug and Substance Abuse, Cardiovascular, Cancer (Breast, cervix, prostate, throat, stomach, ovary and skin), Road Traffic Accidents, Domestic/Occupational Injuries, and Mental Health.

2.3 Advantages of Measurement of Blood Pressure outside the Clinic

Clinical measurement of BP is important because that is where most patients find out that they have raised BP and therefore receive medication accordingly. However, it is not enough only to measure ones BP in a clinical setting (Kaczorowski *et al.,* 2012). According to an article by Kaczorowski *et al.,* (2012) it is important to know how a patient's BP ranges in different situations such as during physical activities, sleep or stressful situations, for better treatment and management of the condition.

Some of the advantages of measuring BP outside of the clinic are: Some patients experience high BP at clinical settings compared to outside of the clinic which is referred to the "white coat" effect. Masked hypertension is one where the patient experiences a high reading of BP at home but a low reading in a clinical setting. This helps reduce unnecessary treatment related to white coat hypertension, it is therefore good for the patient to know what best works for them with respect to measuring BP outside or inside the clinic. Measuring BP on day-to-day basis is important because aside from it being more accurate than a one-off measurement in the clinic, it helps the patients know what makes their BP rise or fall. This also helps in making people aware of their BP levels, whether low, normal or high. It can help patients track how the medicine and lifestyle changes are working for them. Many patients also feel more in control of their condition this way (Krakoff, 2015).

2.4 Factors that Influence Blood Pressure Readings in the Short Term

Unless tested, there are usually neither signs nor symptoms to show if one has high BP or not, but there are some factors that can lower or increase BP temporarily. Over a 24-hour period, BP undergoes variations in response to different activities and stimuli (Mancia, 2012). BP is usually lowest during sleep, rises during the day and starts to fall again in the afternoon. The factors that influence increase of BP in the short term are; caffeine, smoking tobacco, exercise, emotional responses, pain, dehydration, talking and certain medications. The factors that influence decrease of BP in the short term are; warm temperatures (an example is a hot shower), change in posture and also certain medications.

2.5 The Role of Smartwatches in Blood Pressure Management

2.5.1 An Overview of Smartwatches

A smartwatch is defined as a mobile device with a touchscreen display, designed to be worn on the wrist. Different smartwatches come with a variety of diverse features and capabilities and many of them are Bluetooth-capable. They are equipped with different sensors, algorithms, and accompanying mobile applications (Henriksen *et al*., 2018). Aside from touchscreen displays, some can support notifications from the connected phones, others have Global Positioning System (GPS) services, media management, and others can track personal activities (Silbert, 2019). Smartwatches can be used for self-monitoring purposes for example in diabetes or blood pressure.

2.5.2 Studies that used Smart watches in Measuring Blood Pressure

In a study done by Yen & Huang, 2022, the pilot randomized controlled trial explored the effectiveness of commercial smartwatches equipped with a blood pressure-monitoring feature. The study involved participants who were randomly assigned to either a smartwatch group or a control group. The smartwatch group utilized the blood pressure-monitoring feature of the smartwatch, while the control group did not. The researchers evaluated the accuracy of the smartwatch measurements by comparing them with standard sphygmomanometer readings. The findings indicated that the smartwatches provided reasonably accurate blood pressure measurements, suggesting their potential as a convenient and non-invasive tool for monitoring blood pressure.

In a study done by Yen, 2021, the researchers investigated the impact of smart wearable devices as a psychological intervention for promoting a healthy lifestyle and improving quality of life. Participants were randomly assigned to either a smart wearable device group or a control group. The smart wearable device group used the devices to track and monitor various aspects of their lifestyle, such as physical activity and sleep patterns. The study assessed the effects of the intervention on participants' psychological well-being, lifestyle behaviors, and quality of life. The results indicated that the use of smart wearable devices had a positive impact, leading to improvements in psychological well-being, increased physical activity, and enhanced quality of life among the participants.

These studies show how the portability of BP watches and smartwatches that measure BP help patients to comfortably measure BP over a wide range of environments, not only at home. The watches help make it easy to track and record BP readings. Measurement using smartwatches also helps in eliminating the “white-coat effect”. The patient can also measure their BP during different activities in the course of the day providing more insight to them and also to their doctors of what causes their BP to vary. This helps the doctor make better decisions for treatment and the patient to better manage their condition (Yazdi, 2019).

2.5.3 Types of Smartwatches that Can Be Used to Measure Blood Pressure

There are several techniques to measure blood pressure using smartwatches, one like the Omron’s Heart Guide places a thin airbag that sits under the watch band, tightening gradually around the wrist to read BP (Muoio, 2018). This requires one to raise the arm to the heart. The shortcoming of this smartwatch is that it is relatively expensive at $499.00 according to Omron Healthcare.

The VivoWatch BP by Asus uses a combination of EKG (electrocardiogram) and PPG (photoplethysmography) sensors to achieve an accurate reading of the wearer’s BP and heart rate. The patient has to press a finger on the photoplethysmography (PPG) sensor at the top and hold it there for 15 to 20 seconds to get the measurements (Reilly, 2018). During the data collection phase of this research, a relatively fairly priced and easy to use smartwatch for everyday data collection of the patients’ blood pressure was used; namely Smart ECG Temperature Bracelet.

2.6 Applications for Monitoring Blood Pressure

There are many applications that help monitor and control blood pressure. Some of them are; SmartBP, iBP Blood Pressure and Blood Pressure Companion to name a few. These applications monitor the dynamics of BP and help record an individual’s systolic and diastolic pressure, time of measurement, pulse, weight and pulse pressure (Doyle, 2019). Some have reminders to help the patient measure BP, take medication and schedule doctors’ appointments. They also provide data visually through a log, chart, and histogram.

These applications are good at the tasks they perform and can help the patient and the doctor make correct schedules for taking medications and their dosage, but they do not provide more information on how one can manage this condition. HeartHabit, on the other hand, is a blood pressure application for iPhone users that uses artificial intelligence to help patients well manage the condition (Nichols, 2017). It helps those with elevated heart risks, patients that have experienced previous heart attacks or stroke. The application integrates with Apple Watch or Fit bit for secure sharing and portability. These are relatively expensive and presently, the application is only accessible for iPhone users.

2.7 The Use of Machine Learning in Blood Pressure Management

In recent years, researchers began the use of machine learning (ML) algorithms that help patients with hypertension. Different studies and findings use ML algorithms for hypertension in different ways.

2.7.1 Machine Learning for Blood Pressure Prediction

An example is a study done by Nimmala *et al.,* (2018), the age, obesity and cholesterol level of a person was used to predict whether they were prone to have high blood pressure or not (Nimmala *et al.,* 2018). They use data mining classification technique, a decision tree-based J48 algorithm. They proved with an 88.45% accuracy, whether a patient is prone to high BP.

Using two neural network algorithms, back-propagation and radial basis function, an experiment done by Kwong *et al.,* (2016) found that the artificial neural networks (ANNs) were able to predict systolic blood pressure with an accuracy of over 90%. “Artificial Neural Networks (ANNs) are algorithms inspired by function of the human brain. ANNs are usually presented as systems of interconnected "neurons" divided into a few layers. Each neuron can compute values from inputs and is capable of machine learning" (Pytel *et al.*, 2015). The variables used were the age, body mass index (BMI), exercise level, alcohol consumption level, smoking status, stress level, and salt intake level. The new method of predicting BP helped to give early warning to adults who may not get regular BP measurements to know whether it is on an unhealthy level. An isolated measurement of BP is not always very accurate because of daily fluctuations; therefore, their predictor also provided a predicted value as another figure for medical staff to refer to (Kwong *et al.*, 2016).

Figure 1

Structure of an artificial neural network (Kwong et al., 2016)

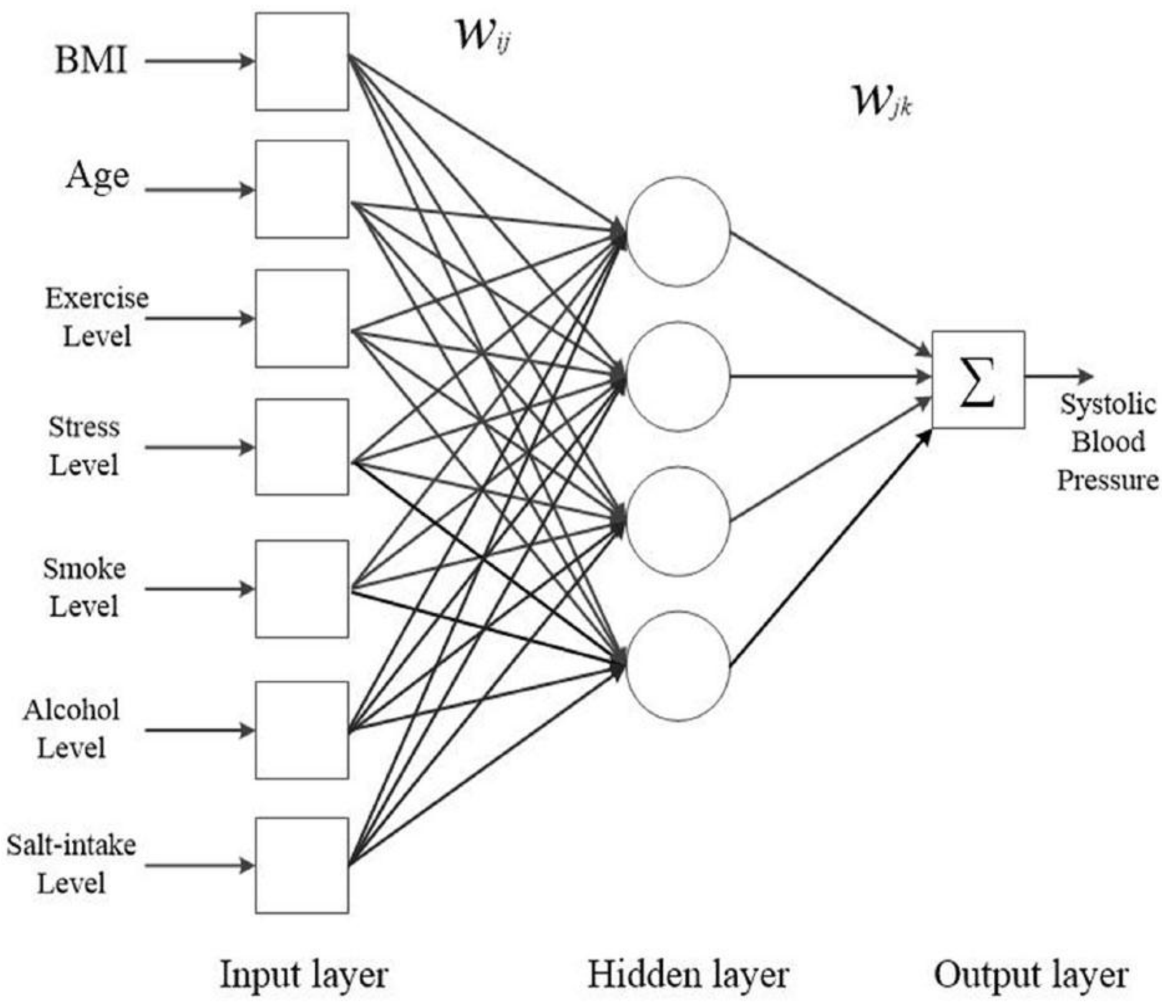


Figure 1 shows the architecture of a typical Back Propagation neural network with seven inputs, one hidden layer with four hidden nodes, and one output value" (Kwong *et al.*, 2016).

A study was done in China using a clinical dataset of 3395 hypertension patients in the city of Shenzhen titled; 3-Year Risk Prediction of Coronary Heart Disease (CHD) in Hypertension Patients (Chen*et al.*, 2017). The data used were the patients age, sex, body mass index (BMI), Systolic Blood Pressure (SBP), heart rate (HR), course of hypertension, Diabetes Mellitus (DM), Hyperlipidaemia (HLP), Chronic Kidney Disease (CKD), Emotional or Mental Disorders (EMD, including anxiety, stress, depression, etc.), and Sleep Disorders (SlpD). Chen *et al*., (2017) found that "traditional risk factors such as age, DM, HLP, CKD for CHD still maintain strong associations with CHD events in hypertension population" (Chen *et al*., 2017). The results also showed that nontraditional risk factors EMD and SlpD had a negative impact on the development of CHD in hypertension patients. They stated that the risk prediction model would help doctors be more accurate in diagnosis and treatment. The study also gives insight to which type of data is associated with high BP.

Golino, et al., (2014), did a study on Predicting Increased Blood Pressure Using Machine Learning. The main problem was obesity in relation to hypertension. The objective was to predict increased blood pressure by body mass index (BMI), waist circumference (WC) and hip circumference (HC), and waist hip ratio (WHR) using machine learning. Classification and regression tree (CART) was used. The results showed that compared to traditional logistic regression analysis, the classification trees produced a much better prediction, with higher pseudo-$R2$, sensibility, and specificity for both men and women (Golino *et al.*, 2014).

Using machine learning to predict hypertension from a clinical dataset, LaFreniere *et al.,* (2016) used a three-layer artificial neural network using backpropagation learning algorithm in order to predict the presence of hypertension in susceptible populations. This was done in order to try to remove the potential for human error. The data used was patients' health conditions, medical records, and demographics**.** Canadian Primary Care Sentinel Surveillance Network (CPCSSN) data set was used. They found that the confusion matrix demonstrates that the network can predict both the presence and absence of hypertension across patient population with about 82% accuracy (LaFreniere *et al.,* 2016).

2.7.2 Machine Learning used for Identification and Diagnosis of Blood Pressure

In 2018, a study was done by Patnaik *et al.*, (2018) to predict the occurrence of essential hypertension using annual health records. The main objective was to predict if an individual will have hypertension in the next year. Data was collected from Korean National Health Insurance Corporation (NHIC) data, containing Electronic medical records. Classifiers; Naive Bayes classifier, Support vector machine (SVM), Logistic regression, Random Forest, Multi-layer Perceptron were used for the study. The results showed that the accuracy range was 66% to 81%, SVM being the most accurate model (Patnaik-Chandran *et al.,* 2018).

Liu *et al.,* (2018) did a study on Identification of Hypertension by Mining Class Association Rules from Multi-Dimensional Features in 2018. Their objective was a class association rules-based method to identify hypertension. This was to utilize the relationship existing in multi-dimensional features to characterize hypertension pattern more effectively, in order to improve the identification performance. Class association rules-based classifier (CAR-Classifier), overlapping sliding window algorithm, LibSVM, Decision Tree, and Naive Bayes used as baseline methods. 128 subjects were recruited, 67 of them were healthy (35 males, 32 females, 53.2±9.2 years), and the rest were hypertension patients (33 males, 28 females, 55.6±7.9 years) which were diagnosed by an experienced physician. Ballistocardiogram (BCG) signal, and Heart Rate Variation (HRV) were used. The researchers found that the relationship among extracted features could be fully exploited, based on which the proposed method identified hypertension with higher performance. CARs were of high consistency, high helpfulness and high usability, which means that the mined CARs could be used as a reference for analyzing hypertension condition in-depth (Liu *et al.*, 2018).

A study done by Nohria (2017) shows how adaptive neuro-fuzzy inference system (ANFIS) for diagnosis of hypertension is better than the existing fuzzy expert system. The data used were the age, blood pressure, body mass index, heart rate, diabetes, physical activities and genetics of the patients. ANFIS is a multi-layer system that integrates the neural network and fuzzy system into single framework (Nohria, 2017). A five layered ANFIS architecture for diagnosis of hypertension was used. Hybrid algorithm was used to modifiable the consequent parameters, to build the ANFIS output equivalent with training data. "Overall accuracy, sensitivity, specificity and precision of our proposed system found to be 94.63%, 97.50%, 93.33% and 98.11% respectively" (Nohria, 2017). In conclusion, the study shows that ANFIS outperforms the existing fuzzy model in terms of accuracy, sensitivity, specificity and precision.

The use of artificial neural networks (ANNs) and anthropometric predictors were used in the diagnosis of hypertension by Pytel *et al.*, (2015) using a dataset of 2485 cases of patients from the city of Lodz, Poland. A three-layer ANN was used in the experiment with the number of neurons in input, hidden and output layers was fixed to 50, 30 and 1 respectively (Pytel *et al.*, 2015). Four models were tested using different inputs, 3 inputs were used for model 1; the body mass index (BMI), the waist circumference (WC) and the age, 3 inputs for model 2; BMI, the waist circumference normalized in relation to the BMI (WCBMI) and the age, 4 inputs for model 3; the BMI, the WC, the sex and the age and 4 inputs for model 4; the BMI, the WCBMI, the sex and the age. According to their results, Pytel *et al*., (2015) state that neural networks should be more widely used for the diagnosis of hypertension.

2.7.3 Association Rule Learning to Study the Relation between Variables That Affect High Blood Pressure

Using association rule learning, research was done by Salehnasab *et al.,* (2014) in order to study the relation between variables that affect high blood pressure. Salehnasab *et al*., (2014) used patient data to assess causes affecting non-communicable disease risk factors on disease associated with high blood pressure. The APRIORI algorithm was used and it was observed that employment factors and physical factors appeared in the individuals that had low blood pressure. Obesity, BMI, low green and fruit consumption appeared in people with high blood pressure (Salehnasab *et al.,* 2014).

2.7.4 Machine Learning in Conjunction with Applications for Blood Pressure Prediction

oHealth: A Self-Care Android Application for Senior Citizens with Hypertension was a study done by Ghoshachandra *et al.,* (2017). Their objective was to design and develop a self-care android application for senior citizens with hypertension to guide, track and treat hypertension and to reduce the risk of having other incurrent diseases. Thirteen variables were selected for analysis which were systolic blood pressure, gender, marital status, smoking status, age, weight, height, body mass index (BMI), exercise level, alcohol use, stress level, hypertension level and salt (NaCl) intake level. 25 participants in the age of 60 to 70 years volunteered to use oHealth. Male participants 36% and 64% female participants. BP5 blood pressure monitoring device was used to measure the participants BP. The researchers used Multiclass Logistic Regression algorithm. The results showed that most of the participants were pleased with the oHealth. It let users know their BP, exercise hour and sodium intake in each day, medicine intake, hypertension levels and learned to be aware of or adopt to a healthy life style (Ghoshachandra *et al.*, 2017).

IoT-Fog based Healthcare Framework to Identify and Control Hypertension Attack was developed by Sandeep K. Sood and Isha Mahajan in 2018. Their objective was to continuously generate emergency alerts of blood pressure fluctuation from fog system to hypertensive users on their mobile phones. The data used was Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Total Cholesterol (TC), High Density Lipoprotein (HDL), Low Density Lipoprotein (LDL), Blood Glucose (BG) and Heart Rate (HR). They also included, environmental data, current physical activity, Global Positioning System (GPS) data, behavioral data and dietary data. Back-Propagation Based Artificial Neural Network (ANN) was used to build their predictive model. The results revealed that ANN gave better results than other comparable algorithms in all statistical measures which justified the use of ANN in the proposed system (Sood & Mahajan, 2018).

The studies mentioned show how far predictive ML algorithms have gone in relation to hypertension. This study took a different dimension in terms of how blood pressure was predicted. Using the data that was collected from the section of the future calendar events of the individuals, the study predicted the individual's future short term blood pressure, whether high, normal or low, using their previous BP data, their activities and future calendar events. The predictions, in turn, could be used to warn individuals of future blood pressure fluctuations so that they would take appropriate measurements to prevent the fluctuations. Due to time constraints, the feedback to be given to the individuals was not part of the study.

2.8 Research Gap

A research gap is a problem which has not been answered either aptly or at all in a particular field of study; it is the missing component in the existing research literature (Pradhan *et al.*, 2018). A researcher conducts research in order to bridge the gap.

2.8.1 Drawbacks of Current Methods for Hypertension Management

A number of hypertensive patients only take hypertensive medications and some go an extra mile to make lifestyle changes as instructed by their physicians. But as it turns out by the growing number of patients with hypertension, these methods are not enough for the management of this non-communicable disease. “Improving BP control is thus unquestionably a goal of fundamental importance for cardiovascular prevention worldwide” (Mancia *et al.,* 2017).The drawbacks here are; the low rate of BP control in hypertensive patients are contributed by physicians’ failure to react with appropriate treatment changes to an uncontrolled BP status, the deficiencies of healthcare systems, and low adherence to the prescribed lifestyle and antihypertensive drug prescriptions as shown by years of research (Mancia *et al*., 2017).

2.8.2 Comparison Table for Machine Learning Models

Table 2

A Comparison of Machine Learning Systems that Predict and/or Manage Blood Pressure

Study Title	Authors and Year	ML/AI Approach Used	Key Outcomes	The Gap
i. Predicting Increased Blood Pressure Using Machine Learning	Hudson Fernandes Golino, Liliany Souza de Brito Amaral, Stenio Fernando Pimentel Duarte, Cristiano Mauro Assis Gomes, Telma de Jesus Soares, Luciana Araujo dos Reis, and Joselito Santos (2014)	Classification and regression tree (CART)	Compared to traditional logistic regression analysis, the classification trees produced a much better prediction, with higher pseudo-$R2$, sensibility, and specificity for both men and women.	The focus of the study was on undergraduate students, not the general population.
ii. Use Association Rules To Study The Relation Between Variables That Affect High Blood Pressure	Cirruse Salehnasab, Fuad Jahandideh, Marzieh Ahmadzadeh, Shahram Tahmasebian (2014)	Association Rule Learning, APRIORI algorithm	It was observed that employment factors, physical factors, and smoking was in conjunction with people with low blood pressure. Obesity, BMI, low green and fruit consumption appeared in people with high blood pressure.	The study did not focus on the constant monitoring of an individual's BP.
iii. Anthropometric Predictors and Artificial Neural Networks in the diagnosis of Hypertension	Krzysztof Pytel, Tadeusz Nawarycz, Lidia Ostrowska-Nawarycz and Wojciech Drygas (2015)	Artificial Neural Networks	According to their results, Pytel *et al*., (2015) state that neural networks should be more widely used for the diagnosis of arterial hypertension.	The study focused on Artificial Neural Networks and not on other types of models like Decision Trees.
iv. A Prediction Model Of Blood Pressure For Telemedicine	Enid Wai-Yung Kwong, Hao Wu and Grantham Kwok-Hung Pang (2016)	Two neural network algorithms, backpropagation and radial basis function	According to the experiment, the accuracy of their predictions of systolic blood pressure values exceeded 90%. Results show that ANNs are suitable for modeling and predicting systolic blood pressure.	The study focused on Neural Networks and not on other types of models like Decision Trees.
v. Using Machine Learning To Predict Hypertension From A Clinical Dataset	Daniel LaFreniere, Farhana Zulkernine, David Barber and Ken Martin (2016)	A three-layer artificial neural network using backpropagation learning algorithm	The confusion matrix demonstrates that the network can predict both the presence and absence of hypertension across patient population with about 82% accuracy.	The data used did not include the users' moods and activity data.
vi. 3-Year Risk Prediction of	Runge Chen, Yujie Yang, Fen	A logistic regression	The results also showed that nontraditional risk	The study did not focus on the

	Coronary Heart Disease in Hypertension Patients: A Preliminary Study*	Miao, Yunpeng Cai, Denan Lin, Jing Zheng, and Ye Li (2017)	model	factors EMD and SlpD had a negative impact on the development of CHD in hypertension patients. They stated that the risk prediction model would help doctors be more accurate in diagnosis and treatment.	constant monitoring of an individual's BP.
vii.	oHealth: A Self-Care Android Application for Senior Citizens with Hypertension	Phukkhapan Ghoshachandra, Chanathip Limkriengkrai Pattarapong Wimonsakcharoe nand Songsri Tangsripairoj (2017)	Multiclass Logistic Regression algorithm	It let users know their BP, exercise hour and sodium intake in each day, medicine intake, hypertension levels and learn to be aware or adopt to a healthy life style.	The use of BP5, a blood pressure monitoring device that is relatively expensive and not as portable as a smartwatch.
viii.	Comparative study of Adaptive neuro-fuzzy and fuzzy inference system for diagnosis of hypertension	Rimpy Nohria (2017)	Adaptive neuro-fuzzy inference system (ANFIS)	94.63% accuracy was obtained from the experiments made on record collected from physician. They obtained sensitivity 97.50%, specificity 93.33% and precision 98.11% for diagnosis of hypertension.	The study did not focus on the constant monitoring of an individual's BP.
ix.	Identification of Hypertension by Mining Class Association Rules from Multi-Dimensional Features	Fan Liu, Xingshe Zhou, Zhu Wang, Tianben Wang and Yanchun Zhang (2018)	Class association rules-based classifier (CAR-Classifier), overlapping sliding window algorithm, LibSVM, Decision Tree, and Naive Bayes used as baseline methods	CARs were of high consistency, high helpfulness and high usability, which means that the mined CARs could be used as a reference for analyzing hypertension condition in-depth.	The Ballistocardiogra m (BCG) signal was only collected in a home environment, and therefore was not as flexible as a smartwatch.
x.	Predict The Occurrence Of Essential Hypertension Using Annual Health Records	Renuka Patnaik, Mahesh Chandran, Seung-Cheol Lee, Anurag Gupta, Chansoo Kim and Changsoo Kim (2018)	Classifiers; Naive Bayes classifier, Support vector machine (SVM), Logistic regression, Random Forest, Multi-layer Perceptron	The results showed that the accuracy range was 66% to 81%, SVM being the most accurate model.	The data used did not include the users' moods and activity data.
xi.	Predicting High Blood Pressure Using Decision Tree-Based Algorithm	Satyanarayana Nimmala, Y. Amadevi, Srinivas Naik Nenavath and Ramalingaswam y Cheruku (2018)	A decision tree-based J48 algorithm	They proved with an 88.45% accuracy, whether a patient is prone to high BP.	The data used did not include the users' moods and activity data.

Source: Researcher (2023)

The current solutions in this field with respect to ML have mostly been used to predict if a person will be susceptible to high BP in the future, find out which variables contribute to a high BP or diagnose high BP and its complications. Some studies predict whether individuals are likely to have a heart attack or other conditions that may come as a result of hypertension. Based on the drawbacks outlined in the introductory part of section 2.8.1, the research gap identified pertains to the insufficient understanding among individuals regarding their blood pressure dynamics, specifically in relation to temporal fluctuations resulting from daily activities, emotional states, and various other influencing factors. This study made future short-term predictions of an individual's BP using data collected from the individuals inclusive of the activities, and an ML model. Smartwatches were used for the purposes of portability. A smartphone application was also used for the purposes of data collection. The BP predictions can help individuals have more control over activities and moods and therefore help in the regulation ofBP before it gets to dangerous levels.

2.9 Theoretical Framework

Theories are formulated to explain, predict, and understand phenomena. In many cases, they exist to challenge and extend existing knowledge within the limits of critical bounding assumptions. Thus, a theoretical framework is the structure that can support a theory of a research study (University of Southern California, 2020). The following subsection will discuss how machine learning works and what gradient boosting regression is.

2.9.1 Machine Learning

Machine learning is a branch of computer science that focuses on creating algorithms and models that allow computers to learn and make predictions or judgments without being explicitly programmed. ML is based on the concepts of pattern recognition and statistical inference. It is based on the utilization of training data to train models, which then generalize patterns and correlations from the data to make correct predictions or choices on unseen or future data. Models learn by iteratively modifying their parameters depending on training data, with the goal of minimizing errors or optimizing performance measures(Zhou, 2021).There are three types of machine learning models: supervised learning, unsupervised learning, and reinforcement learning.

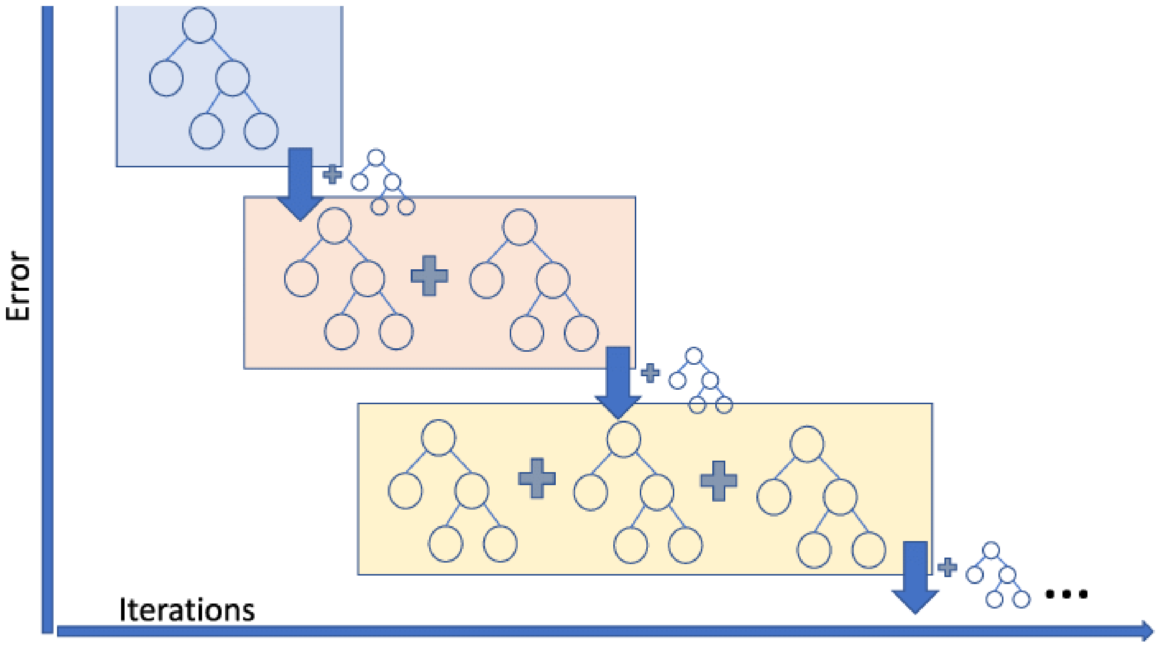
Supervised learning models learn from labelled training data, where the input data is associated with corresponding output labels. These models seek to generalize the mapping between inputs and outputs in order to predict or classify new, previously unknown data(Zhou, 2021). Decision trees, random forests, support vector machines (SVM), and deep neural networks are examples of supervised learning techniques. Unsupervised learning models, on the other hand, work with unlabelled data to find underlying patterns or structures. These models lack predetermined output labels and instead concentrate on clustering comparable data points or reducing data dimensionality(Zhou, 2021).Unsupervised learning algorithms that are commonly used include k-means clustering, hierarchical clustering, and principal component analysis (PCA).Reinforcement learning models, on the other hand, learn through interaction with an environment.

These models seek the optimum actions in order to maximize a cumulative reward signal. Reinforcement learning algorithms develop optimal policies through trial and error by experimenting with various behaviours and getting feedback from the environment(Zhou, 2021). Q-learning, Deep Q-Networks (DQN), and Proximal Policy Optimization (PPO) are examples of notable reinforcement learning algorithms.

Every type of machine learning model possesses a unique array of strengths and weaknesses, and the selection of the most suitable model is contingent upon factors such as the problem's characteristics, the available dataset, and the desired outcome. A good example of a machine learning algorithm is the Gradient Boosting Regressor. Gradient boosting is a variant of ensemble methods that combines multiple weak models to improve overall performance. Boosting is a machine learning technique for combining multiple simple models into a single composite model. Boosting is also known as an additive model because it adds single simple models while maintaining the integrity of the model's current trees. The term gradient in gradient boosting refers to the algorithm's use of gradient descent to minimize loss (Dhiraj, 2019). Gradient boosting can predict challenging targets by combining multiple weak models (Masui, 2022).

Figure 2

Gradient Boosting Decision Trees (Baturynska & Martinsen, 2021).



These are the steps that GBR uses according to Brownlee (2020):

i. A loss function that must be optimized.

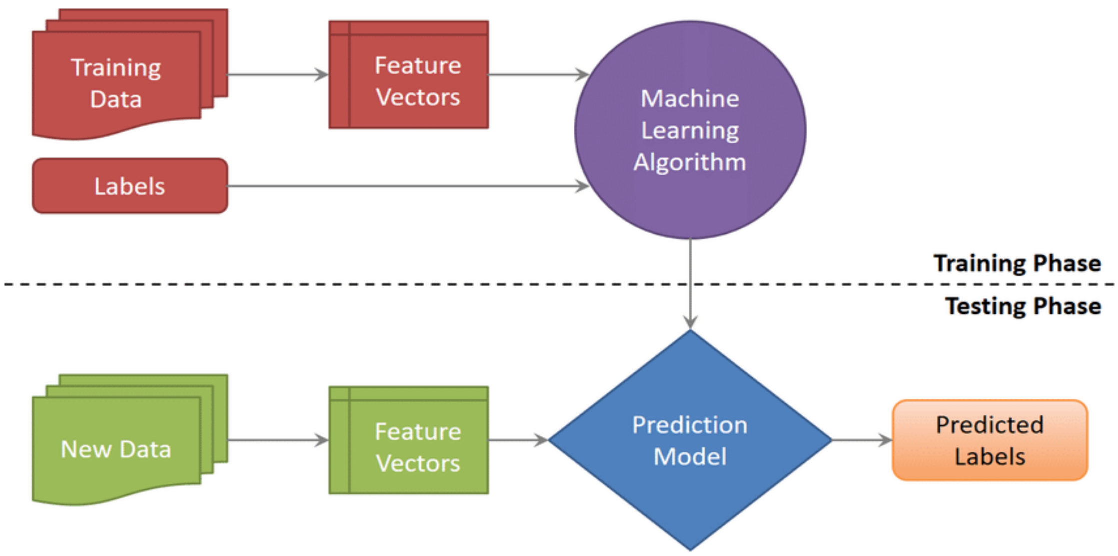
ii. Predictions made using weak learners.

iii. To minimize the loss function, an additive model is used to add weak learners.

Gradient Boosting is used for regression when it is used to predict a continuous value such as age, weight, or SBP. This is different from using linear regression. Gradient Boosting can also be used for classification; an example of Gradient Boosting Classification is when it is used to predict whether or not a patient has hypertension or not. In this study, Gradient Boosting Regressor was used because of the prediction of the continuous variables of BP. The GBR algorithm is among the most powerful models in machine learning. It is used to reduce the bias error of a model. Other advantages of GBR are; it provides higher flexibility, has better accuracy, requires less pre-processing and handles missing data on its own (Dhiraj, 2019). Guo (2017) outlines a comprehensive seven-step framework for machine learning.

Figure 3

Typical machine learning procedure (Ezzat, 2018)



Note: The Seven Steps of Machine Learning (Guo, 2017)

Step 1: Gathering Data

The quality and quantity of data will determine how good the predictive model will be. The data collected will then be tabulated, this is called Training Data (Guo, 2017). For the study, examples of the data that were collected are BP readings and activity data.

Step 2: Data Preparation

During data preparation, data is loaded into a suitable platform, an example is Microsoft Power Query for Excel, in which data will then be prepared for the machine learning training. The data should be randomized so as not to affect how the model will learn. Data visualization is also done in this step, this helps the researcher see relationships between different variables, as well as data imbalances that are present. The process of data visualization offers valuable insights into the prepared data, which may otherwise go unnoticed when presented solely in tabular format. Data manipulation is also done in this stage and involves modifying, transforming, or reorganizing raw data to prepare it for analysis. Additionally, data normalization, which standardizes data to enable equitable comparisons and enhance machine learning algorithm performance, is also conducted during this stage. While in the preparation section, data will be split in three parts, the first part, which is the majority of the dataset, is the training dataset 70%, to train the model, the second part is the validation dataset 15%, to evaluate the model and the test dataset another 15% to examine the trained model for its performance.

Step 3: Choosing a Model

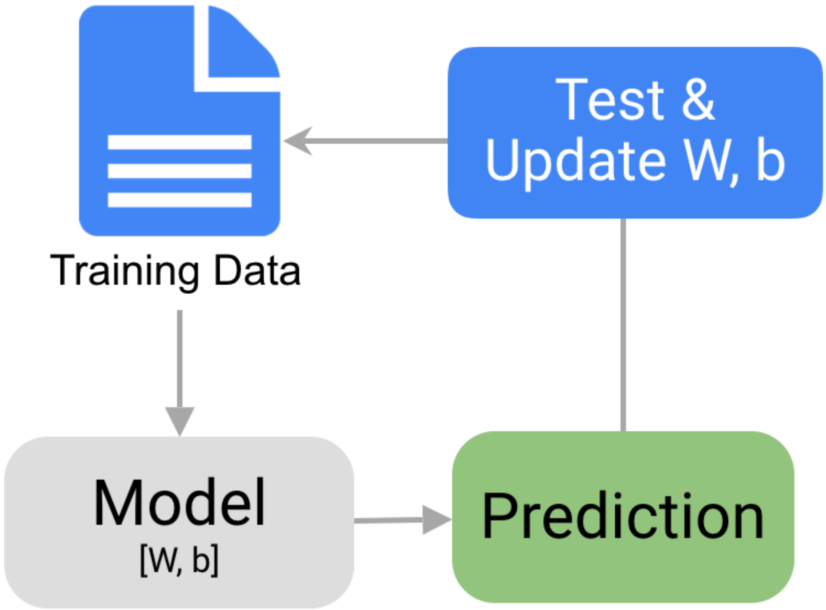
Many models have been developed by scientists and researchers over the years. Choosing the right model is crucial so as to get correct predictions. Gradient Boosting Regression was used to develop the model. Sci-Kit Learn was used as a framework to develop the machine learning model.

Step 4: Training

Training is where the data is used to incrementally improve the model's ability to predict. In machine learning, we have weights (W) and biases (b). "The training process involves initializing some random values for W and b and attempting to predict the output with those values" (Guo, 2017). To have more correct predictions, the values W and b have to be adjusted accordingly. A cycle of updating the weights and biases is called one training step.

Figure 4

Training the model (Guo, 2017).



Step 5: Model Evaluation

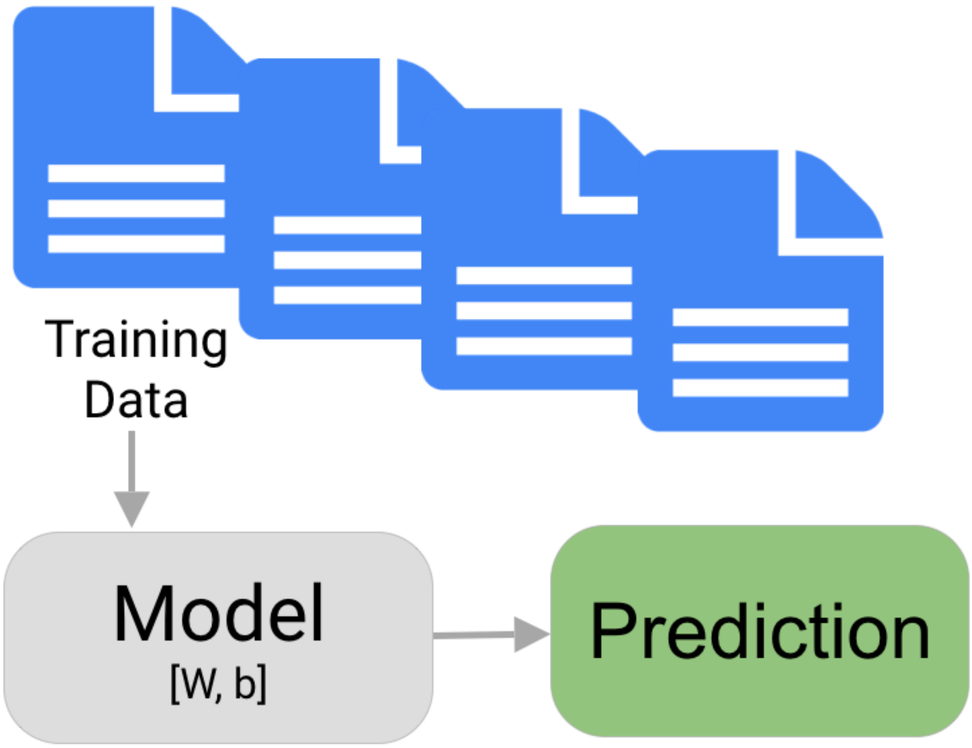
Evaluation is the process of testing whether the model trained is good enough. This allows for testing of the model against data that has never been used before for training. It involves using the remainder of the data not used for training. This will show how the model might perform in the real world.

Step 6: Parameter Tuning

Parameter tuning is done in order to further improve the training of the model. This helps correct assumptions made in earlier training of the model. Guo (2017) gives an example of how many times the training dataset is run through during training. The model can be shown the dataset multiple times instead of ones.

Figure 5

Parameter Tuning (Guo, 2017)



This is an experimental process that depends on the dataset provided, the model used and the training process.

Step 7: Prediction

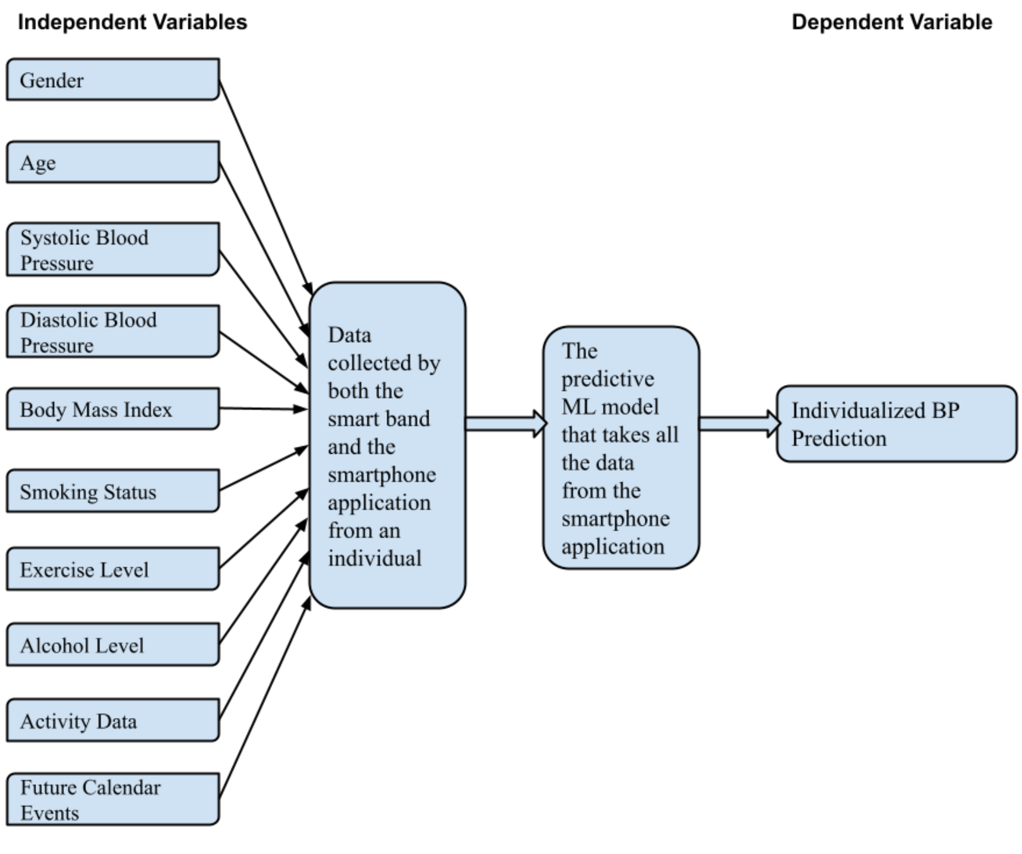
According to Guo (2017), machine learning is "using data to answer questions" (Guo, 2017), this is where the model was used to provide inferences about the BP of the individuals.

2.10 Conceptual Framework

The causes of hypertension are often not known though there are risk factors associated with the increase of BP. Some of these factors are; age, race, weight, alcohol use, tobacco use, gender and existing health conditions such as chronic kidney disease (MacGill, 2018). Some of the ways in which a patient can monitor and control their BP is by measuring BP at home, taking medication if recommended by the doctor and making lifestyle changes such as eating healthy, reducing sodium intake, exercising, reducing the amount of alcohol intake, quitting smoking and reducing stress. These are some of the data that the researcher collected in order to create a predictive machine learning model. In this case, BP prediction is the dependent variable, some of the independent variables are; BP readings, exercise and alcohol intake.

Figure 6

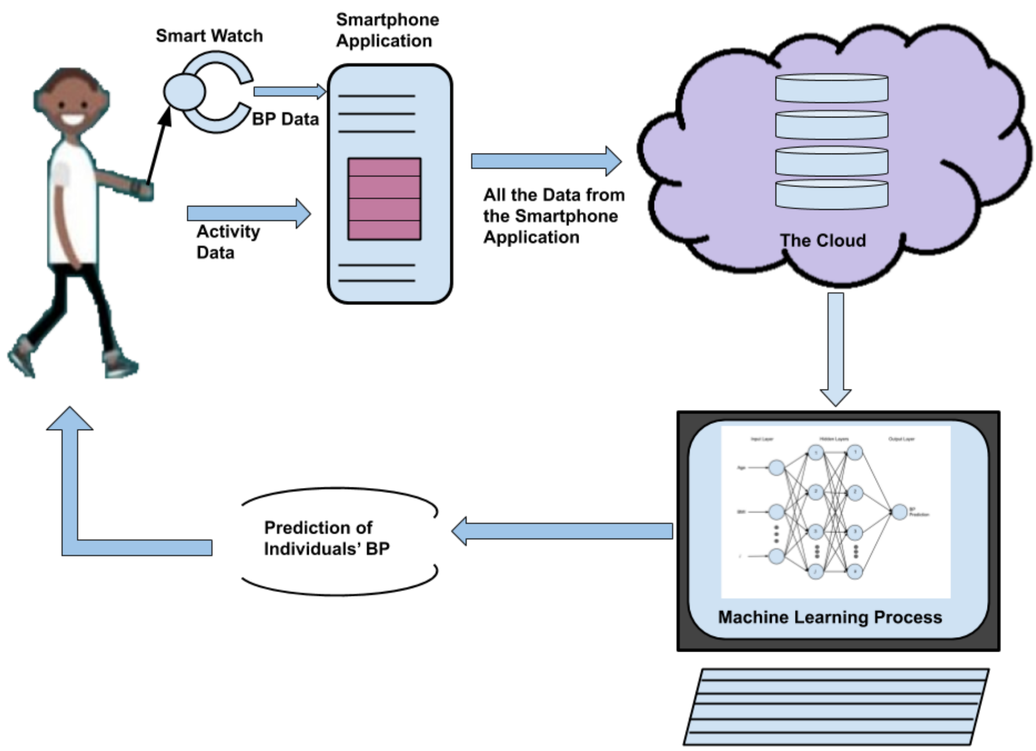
Process Flow for Blood Pressure Prediction



The independent variables that were collected are the BP readings, exercise, tobacco, alcohol levels, activity data, and the future calendar events. This data was collected through the smartwatch and from the individuals input in the smartphone application. The data stored in the smartphone application was then sent to the cloud and then to the predictive machine learning model in order for it to learn and finally predict the individuals' BP fluctuations. The BP predictions the dependent variable in this case.

Figure 7

The Proposed System Architecture for the Collection of Data and Prediction of an Individual's BP



The smartwatch takes BP readings from the individual, which is then transmitted to the smartphone using Bluetooth technology. The individual also provides data such as the activity data, alcohol intake, tobacco intake and future calendar events to the smartphone application. The data is then sent to the cloud and then used by the machine learning model to be able to learn and predict the individual's BP.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter discusses the research design, the location of the study, population of the study, sampling procedure and sample size, the instrumentation with its validity and reliability, model development, data collection and analysis, evaluation and ethical consideration.

3.2 Research Paradigm

A research paradigm is a broad framework or worldview that influences a researcher's approach to examining a specific phenomenon. It includes the researcher's views, assumptions, and philosophical foundations, which impact their view of how information is acquired and perceived. Among the most common research paradigms are positivism, interpretivism, critical theory, and constructivism(Abdul Rehman & Alharthi, 2016). The research paradigm serves as the primary lens through which the researcher understands and conducts their research.

The research paradigm for the study followed the positivist paradigm. Positivists assert the existence of a measurable and understandable reality. Positivist studies often commence with the formulation of empirical hypotheses, subsequently evaluated through data collection and analysis. An objective, scientific approach is adopted to examine quantitative relationships between variables, rather than seeking qualitative explanations. Positivists maintain that findings from one study can be applied to comparable scenarios(Ulz, 2023).

3.3 Research Design

The methodology that was used for the study is design science and experimental methods. Design science involves the creation and evaluation of Information Technology artefacts that are intended “to solve observed problems, to make research contributions, to evaluate the designs, and to communicate the results to appropriate audiences.” (Peffers *et al*., 2007). The objective of design science research is to generate actionable knowledge that professionals in the respective discipline can apply to design effective solutions for the challenges encountered within their field. Experimental methods involve “manipulating one variable to determine if changes in one variable cause changes in another variable. This method relies on controlled methods, random assignment and the manipulation of variables to test a hypothesis” (Cherry, 2019).

These methods were relevant to the study because of its nature which involves the creation of an application which collects data from the participants. The data was then used by the machine learning model which was created, in order to predict an individual’s BP. This system’s performance was then validated.

3.4 Location of the Study

This research took place in Uasin Gishu County.The county has an estimated population of 1,163,186 as of the census done in 2019 (Kenya National Bureau of Statistics, 2019). The population selected from this area was able to provide the data needed for the study. The study was done in this county because of the presence of Moi Teaching and Referral Hospital (MTRH) which provided some insight about the prevalence of hypertension in the Rift Valley region as it serves this region including neighbouring counties like Uganda. It was a starting point; further studies will constitute other counties in Kenya.

3.5 Population of the Study

The population for the study comprised of people who do not have known hypertension. The target population therefore were individuals whose hypertensive condition is not known.This is becausehigh BP has no obvious symptoms to show that something is wrong, thus also known as the “silent killer”.The inclusion criteria for the target population was individuals who have no known hypertension and are 35 years and above; this is because according to the World Health Organisation (WHO) (*Hypertension*, 2023), approximately 1.28 billion adults between the ages of 30 and 79 have hypertension around the globe. The exclusion criteria for the target population was individuals who have known hypertension and are under high BP medication.

3.6 Sampling Procedure

The sampling procedure for this study was quota sampling. This is a sampling method of gathering representative data from a group. It is a process that takes a tailored sample that is in proportion to some characteristic of a population. It ensures that the sample group represents certain characteristics of the population chosen by the researcher (Aprameya, 2016). Quota sampling uses non-random sampling methods to gather data from one stratum until the required quota is fulfilled. The researcher took a sample of the target population from each constituency in Uasin Gishu County. This method was picked for the study because of the type of study and limited resources.

In the case of this study, each participant was required to give data through the smartwatch and the smartphone application on a daily basis for a period of 2 weeks in order to get the information needed for the BP prediction model.

3.7 Sample Size

According to the quota sampling technique, which uses non-random sampling, the survey population is divided into subgroups. The subgroups are selected with respect to certain known features, traits, or interests (Aprameya, 2016). Mohamed et al., (2018) in their study stated age as one of the factors that contributed to high BP. Therefore, the population of the study included individuals of 35years and above. There were two subgroups according to gender, male and female. For each subset, the individuals selected had no known hypertension condition. In this case, the data needed from the participants was mostly qualitative, which did not require a large sample size. Non-random sampling enables researchers to target a specific group of interest with distinctive qualities or attributes relevant to the research aims. This method allows for in-depth examination and comprehension of a specific population that would not be possible with random sampling.

In this type of sampling, judgement of the researcher is required in choosing the number of participants for the study. The data collection was conducted incrementally, and the sample size was deemed sufficient upon reaching a saturation point. Through this iterative process, data collection continued until no new information emerged, ensuring a comprehensive understanding of the research objectives were achieved. The total number of participants for the study thereforewas45 individuals. The data collected from the individuals ensured the adequacy of the sample size. The male participants were29 which is 65% of the participants, while the female participants were16, which is 35%. This is because the hypertension condition is higher in males compared to females and so is the awareness of the condition which is also low in males compared to females (Mohamed *et al.*, 2018).

3.7 Instrumentation

The data for this study was collected through the smartwatches and the smartphone application provided to the individuals.

The smartwatch specifications;

i. EKG also known as ECG detection for reading the heartrate and blood pressure.
ii. Bluetooth 4.0, which enabled connectivity with the smartphone application.
iii. Rechargeable battery.
iv. Display screen of 0.96 inches.

The specifications for the smartphone;

i. Android Version 4.0 and above.
ii. Bluetooth Version 3.0 and above.
iii. Global Positioning System (GPS) enabled.

The BP was measured via the smartwatch and sent to the smartphone application. The future calendar events were provided by the users using the phone application, and the participant manually entered data such of the activities done during the day. The data collected was then sent to the cloud and used to train the predictive ML algorithm.

3.7.1 Pilot Study

A pilot study was done at Kabarak University. This involved giving the smartwatches and installing the application to the individuals who were willing to participate. This helped in making sure that the phone application and the smartwatches work well. It also helped to verify that the data collected was sufficient for the model to make predictions.

Five participants took part in the pilot study. The smartwatches used for the pilot study were not able to collect data from dark skinned individuals since it only used a photoplethysmography (PPG) sensor. This limited the sample size of the pilot study.

This limitation forced the researcher to look for better smartwatches that could read people of all skin tones for the main study. A smartwatch with both the PPG sensor and the electrocardiogram (ECG) sensor was used for the main study. The application used for the pilot study also informed the main study by adding useful features that made it more convenient for the participants to record data. These features encompass the implementation of a more automated method for recording BP, the inclusion of emoji in the moods section as a supplement to words, and the development of a user interface that is more intuitive in nature.

3.7.2 Validity of the Instrument

Construct validity is used to determine how well a test measures what it is supposed to measure. It evaluates whether a measurement tool really represents the thing we are interested in measuring. "A construct refers to a concept or characteristic that can't be directly observed, but can be measured by observing other indicators that are associated with it" (Middleton, 2020). Construct validity is usually verified by comparing the test to other tests that measure similar qualities to see the correlation between the measurements. In this study, construct validity of the type of data collected from the individuals, which is the BP and the activity data, is ensured because the indicators and measurements are carefully developed based on relevant existing knowledge from other studies done of hypertension.

3.7.3 Reliability of the Instrument

According to Trochim (2006) the test-retest reliability (repeatability)is estimated when the same test is administered on the same sample on two different occasions, two different times T1 and T2. "This approach assumes that there is no substantial change in the construct being measured between the two occasions" (Trochim, 2006). The time difference here is important, this is because a short time gap, yields higher correlation compared toa longer time gap. This was done by measuring the BP of individuals several times within a short period of time to see whether the smartwatch was reliable in its measurements.

Internal Consistency Reliability was also used as a measure of reliability of the instrument. It is the measure of consistency of the results across items. The internal consistency reliability test was done by measuring the accuracy of the smartwatches and comparing them to a doctor's office BP monitoring device several times. The smartwatch, the digital BP machine and the sphygmomanometer were used, and the smartwatch readings were close to the readings of the sphygmomanometer, which shows that it is quite accurate.

3.8 Model Development

"Model development is an iterative process, in which many models are derived, tested and built upon until a model fitting the desired criteria is built" (McGahey& Cameron, 2002).

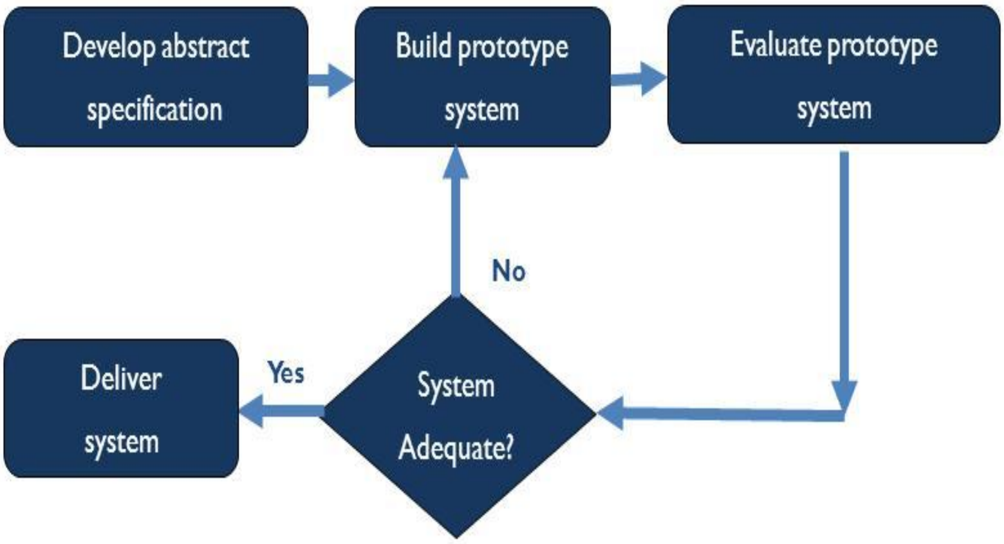
3.7.1 Prototype Development

According to Camburn *et al.,* (2017), a prototype is an artifact that approximates a feature or features of a product, service, or system (Camburn *et al.*, 2017). It is an example that serves as a basis for future models. It is used as a proof-of-concept and usually influences the project's success. A prototype can also be used to demonstrate to users on how the system will look like and how it will work.

Evolutionary prototyping is a development method where the developer first constructs a prototype and hands it to the user. The feedback received from the user will then inform the developer on how to improve the system by adding functionality and/or improvements until the product is satisfactory (Sherrell, 2013). This is the method used in the development of the prototype, as a whole.

Figure 8

Evolutionary prototyping model. Reprinted from CrackMBA, 2012

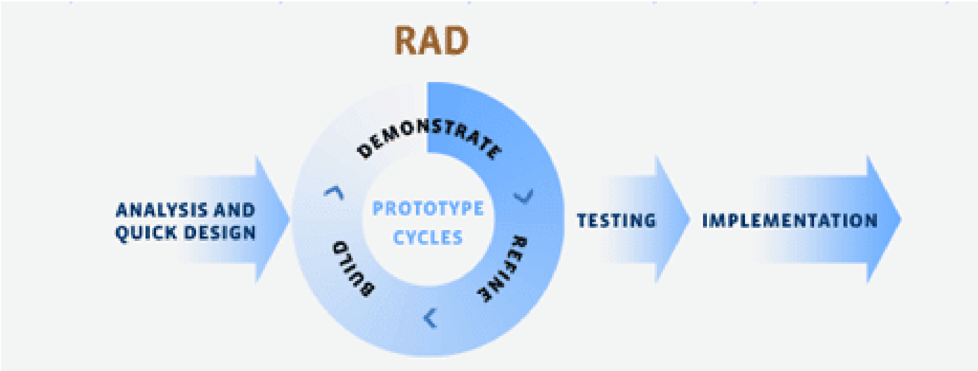


3.7.2 To Develop a Model for Regular Collection of Blood Pressure and Activity Logging, Comprising of a Smartwatch and Mobile Phone Application

For the first objective, which is to develop a prototype comprising of a smartwatch and mobile phone application for the regular collection of blood pressure data and activity logging, Rapid Application Development (RAD) was used. The components of the model are the smartwatch and the mobile application.

Figure 9

Rapid Application Development (RAD). Reprinted from Testing Excellence 2018



This process involves analysis and quick design, a cycle that contains demonstration, refinement and building, what follows after the cycle is testing and finally implementation. The application developed allows for data retrieval from both the smartwatch and the participant.

The data that was collected from each participant included; BP readings, activity data, weight, age, BMI, smoking status, alcohol level and stress levels from the 45 participants. The individuals took the BP and heart rate readings using the smartwatch, the smartphone application was then used to take the rest of the data inclusive of the BP and heart rate readings collected by the smartwatch from the individual via Bluetooth technology.

The tools that were used in order to develop the smartphone application for it to pick specific data from the smartwatch and also from the individual are; the integrated development environment (IDE) from Android Studio. This is the official IDE for Google's Android operating system (Ducrohet *et al.*, 2013). Firebase, which is an application development platform, was used for database purposes. The programming language that was used is Java. Bluetooth Low Energy (BLE) was used for communication between the smartwatch and smartphone.

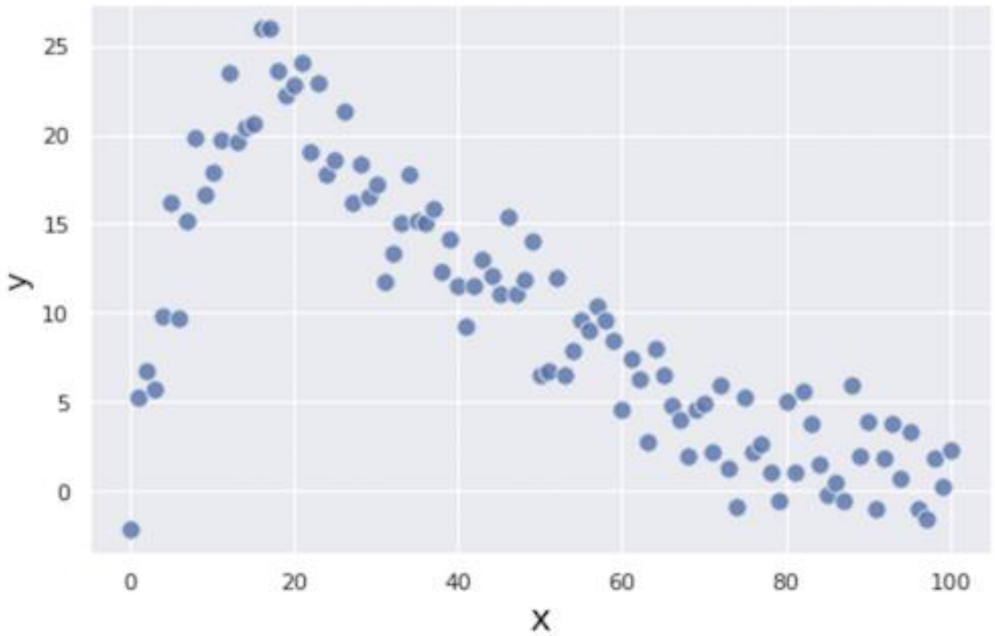
3.7.3 Implementation of the Gradient Boosting Regression Predictive ML Model

For objective two, the implementation of the ML model was done using Iterative and Incremental Development. This involves developing a system through repeated cycles. It allows for learning through both development and use of the system (Ugurlu, 2018).

Masui (2022) illustrates an example along with explanations in order to better understand how GBR works. Gradient boosting regression trees will be illustrated step by step using a sample with a nonlinear relationship between x and y.

Figure 10

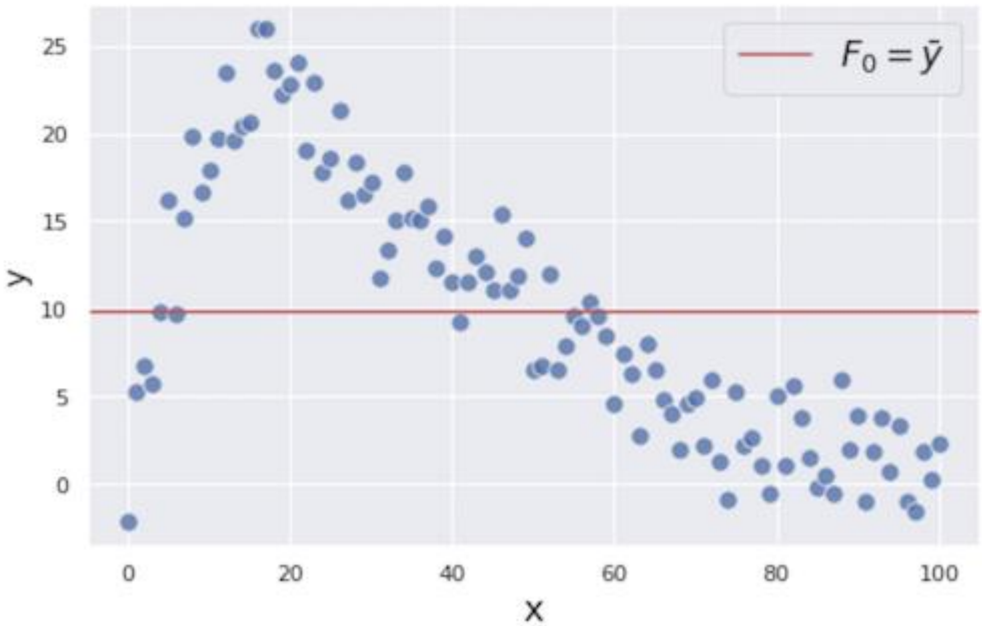
The figure represents a sample for regression problem (Masui, 2022).



The first step is to make a naive prediction about the target y. As an overall average of y, we make the initial prediction F_0.

Figure 11

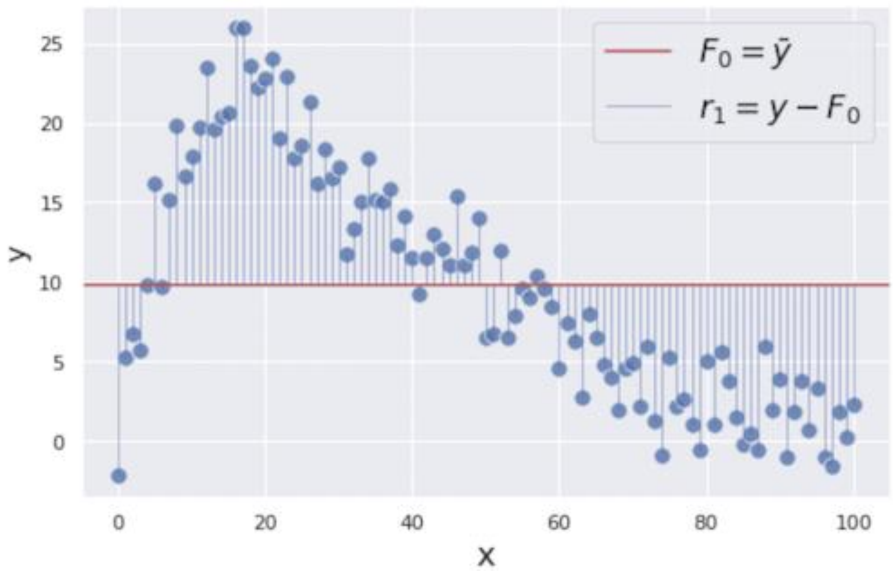
This shows the initial prediction F_0 = mean(y) (Masui, 2022).



In this first instance, the mean has been used for prediction. This is an example of a weak predictive model. To improve the prediction, we will start with the residual's prediction errors or the residuals because that is what needs to reduce to get a better prediction. The residuals r_1 are depicted in the figure below as vertical blue lines (Masui, 2022).

Figure 12

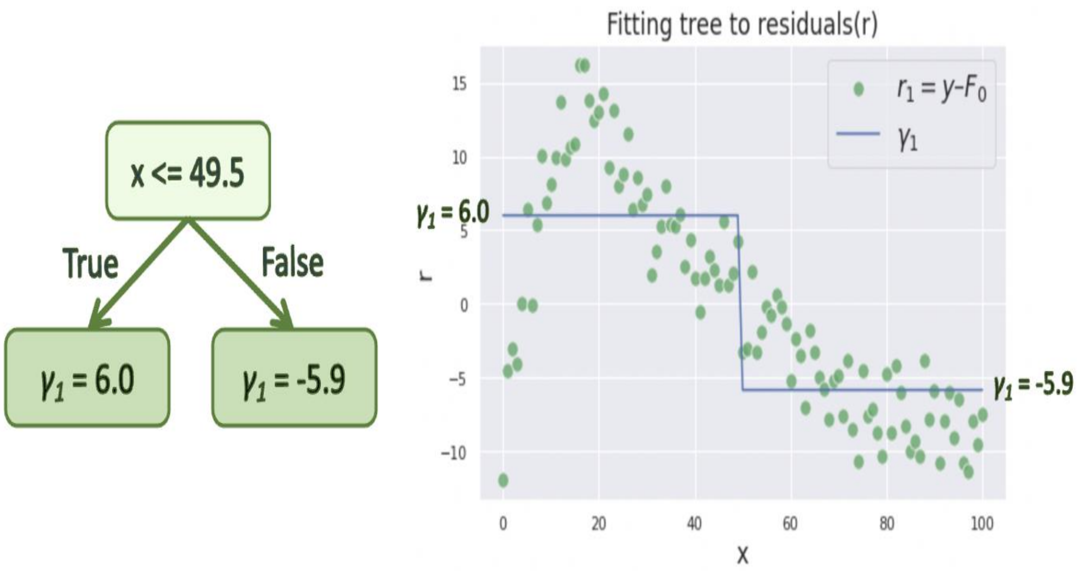
The figure shows the residual r_1 (Masui, 2022).



To minimize these residuals, a regression tree model with x as its feature and the residuals r1 = y mean(y) as its target is built. This is because if some patterns between x and $r_1$can be found by building the additional weak model, they can be used to reduce the residuals. Gradient boosting trees typically have between 8 and 32 terminal nodes, but in this example, simple trees are constructed, each with one split and two terminal nodes. The first tree has been created predicting residuals with two different values y_1 = 6.0, -5.9. The prediction is represented by y, which is gamma.

Figure 13

The figure shows a tree which has predicted 2 residual values (Masui, 2022)



To reduce the residuals, this prediction y_1 is added to the initial prediction F_0. The gradient boosting algorithm does not simply add y to F because it causes the model to overfit to the training data, rather, y is scaled down by the learning rate v, which ranges from 0 to 1, and then added to F.

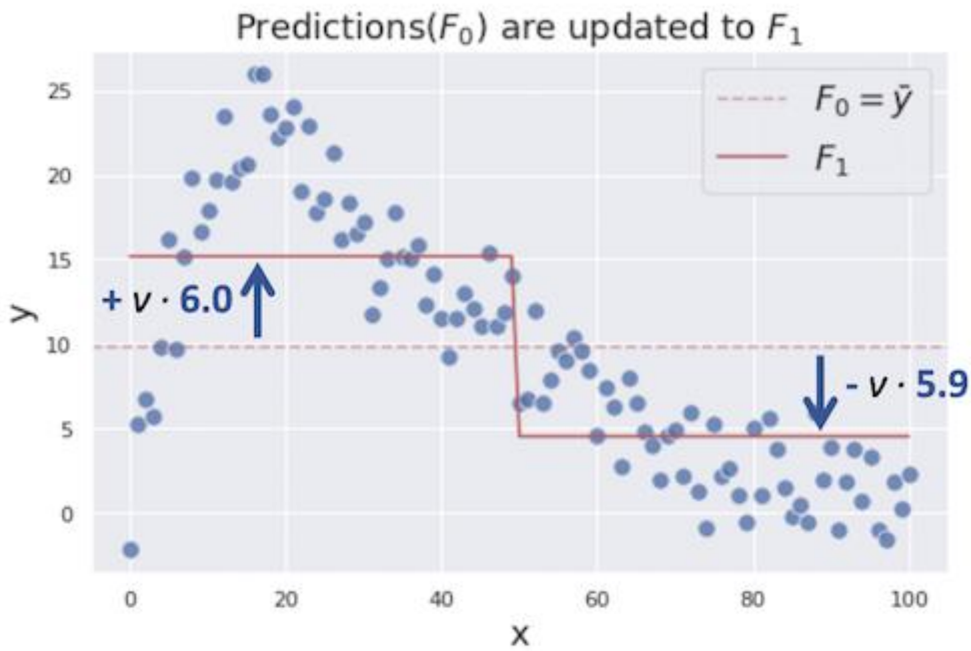
$$F_1 = F_0 + v.y_1$$

To make the optimization process easier to understand, the illustration uses a relatively large learning rate v = 0.9, but it is usually much smaller, such as 0.1. After the predictive model has been updated the combined prediction F_1 is:

$$F_1 = \begin{cases} F_0 + v.6.0 \; if \; x \leq 49.5 \\ F_0 - v.5.9 \; otherwise \end{cases}$$

Figure 14

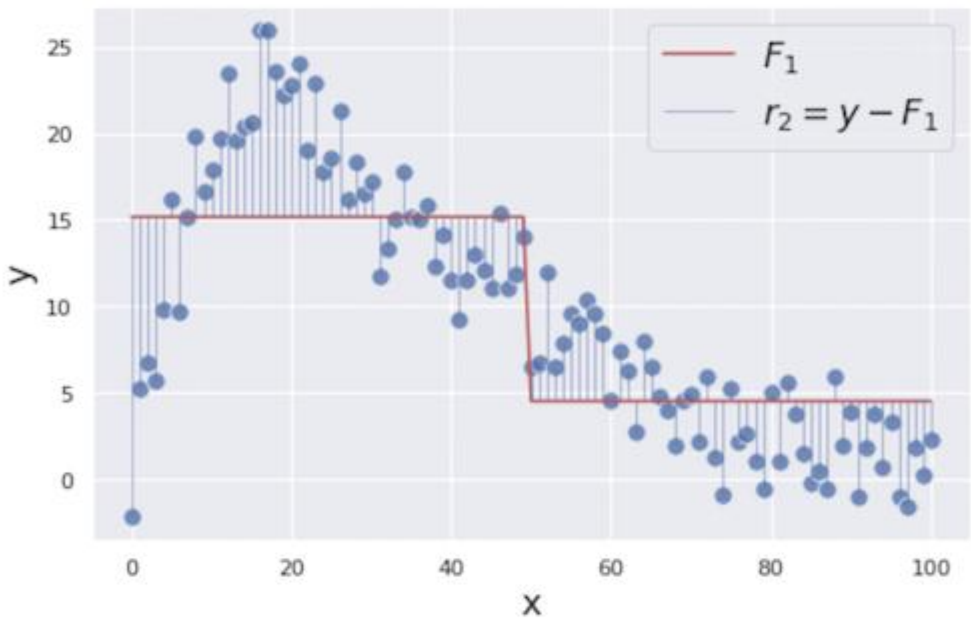
The figure shows the combined prediction (Masui, 2022)



The figure below depicts the updated residuals r_2 .

Figure 15

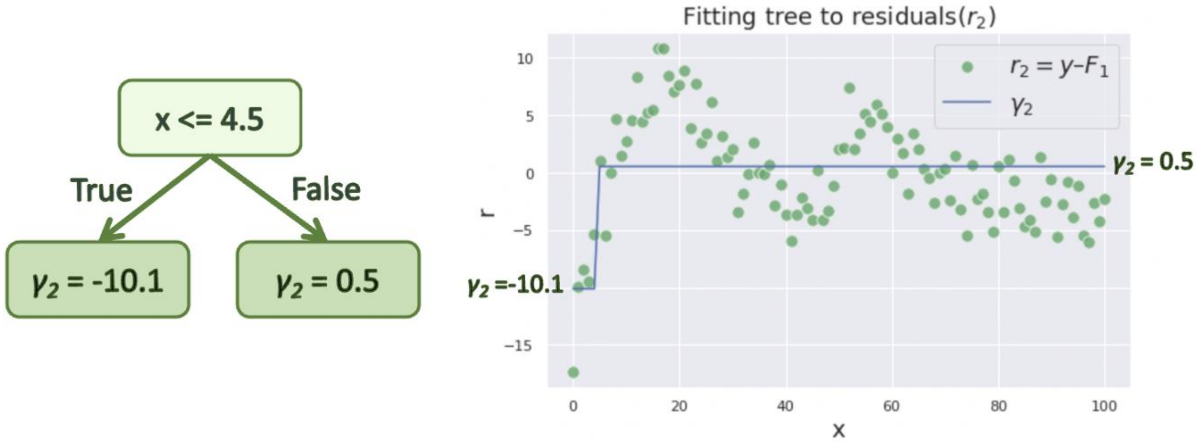
The figure shows the updated residuals r_2(Masui, 2022)



In the following step, another regression tree is built with the same x as the feature and the updated residuals r_2 as the target. The created tree is depicted in the figure below.

Figure 16

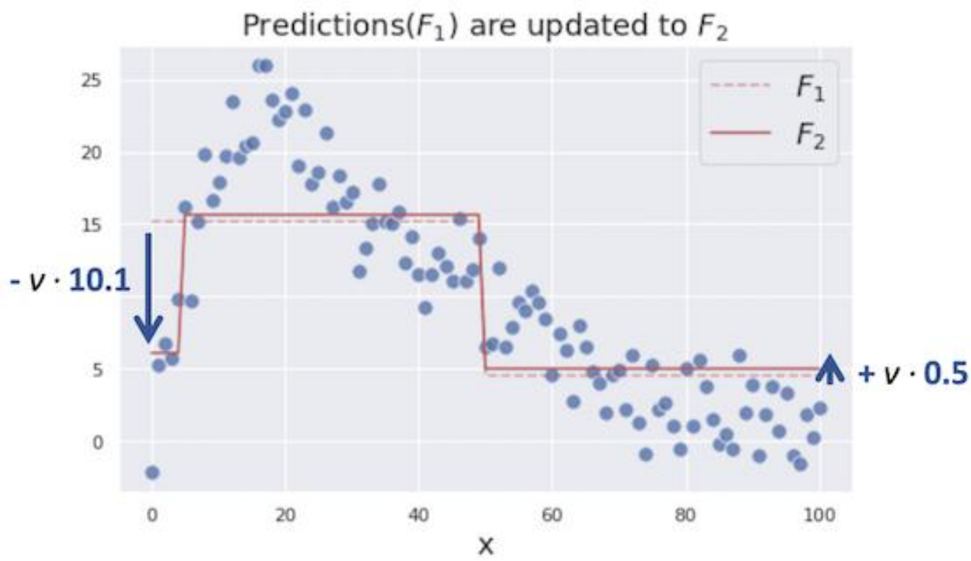
Another tree has been created using x as the feature and the updated residuals r_2 as its target (Masui, 2022)



After that, the previous combined prediction F_1 is updated with the new tree prediction y_2.

Figure 17

Updating the previously combined prediction F_1 with the new tree prediction γ_2 (Masui, 2022).

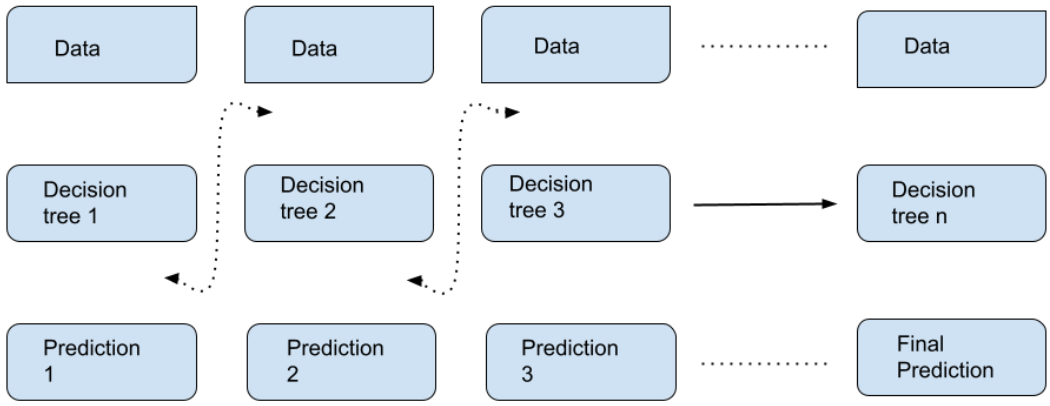


These steps are iterated until the model prediction no longer improves.

Figure 18

The training process of a Gradient Boosting Regressor Model (Researcher, 2022)

The process of Training a Gradient Boosting Regression Model



The steps;

i. The approach that the study uses is supervised training, the first part involves defining the input and target data.

ii. The pre-processed data was then divided into 3; Training dataset, Validation dataset and Test dataset.

iii. Input the training dataset, for training the model.

iv. Make the first prediction.

v. Then compute the pseudo residuals.

vi. Then predict the pseudo residuals.

vii. Make another prediction.

viii. Repeat steps iv, v, vi and vii until minimal error is obtained.

ix. Input the validation dataset. This is for evaluation of the model. It is used to fine-tune the model’s hyper-parameters; an example is the number of estimators; this represents the number of boosting stages that the model performs.

x. The prediction obtained is then compared to the expected label and the error is calculated.

xi. If the minimum accepted value is obtained, the model can then be simulated.

For training and testing the ML model, Scikit Learn, Numpy and Pandas libraries and packages were used and Jupyter Notebook was used as a tool to write machine learning code using Python language.

3.8 Data Collection

The researcher delivered the smartwatches to each participant together with the smartphone application. These are the instruments used to collect the participant's data. The researcher then elaborated to the participants on how to use both the smartwatch and the application. The data collected was then stored in the cloud.

Table 3

Data Collection Table

Independent Variable	Description	Measure
Gender	Constant variable	M, F
Age	Continuous variable	Year
Systolic blood pressure	Continuous variable	mmHg
Diastolic blood pressure	Continuous variable	mmHg
Heart Rate	Continuous variable	beats per minute
Body mass index (BMI)	Continuous variable	kg/m^2
Exercise level	Binary variable	Yes, no
Smoking status	Numerical Variable	
Alcohol level	Binary variable	Yes, no
Sleep Level	Numerical Variable	Hours
Medication	Numerical Variable	
Mood	Categorical variable	
Activity Data	Categorical variable	
Future Calendar Events	Categorical variable	

Source: Researcher (2023)

3.9 Data Analysis

Before inputting data into the ML model, it must go through some steps of data pre-processing and data analysis. The data maybe incomplete, contain duplicates or errors, therefore data cleansing was needed. This process involved identification of null values, inaccurate data, record matching and de-duplication of data. Most of the ML algorithms require perfect data and in a specific format so as to reduce errors. Therefore, Pandas, NumPy and Scikit-learn libraries were used for data transformation, data manipulation and data pre-processing. This includes checking for missing values, categorical data, standardization of data, data normalization and splitting of data. Once the data was free of errors it was analysed and translated into a format that the ML model could understand and then the first set of data, which is the training data, was then inputted for the purposes of training the model.

Data pre-processing involved:

3.9.1 Data Quality Assessment

To determine whether the data collected was adequate for the project, the overall quality, consistency, and relevancy of the data were evaluated. Various tables containing various types of features were gathered and integrated. Before beginning the data cleaning procedure, any obvious data irregularities were addressed.

3.9.2 Data Cleaning

Cleaning data entails one or more of the following steps: correcting, deleting extraneous data, or adding missing values to the data set. MySQL and Pandas were used in this process. The null values were filled, and the null values that couldn't be filled in were removed. There were several columns in the study that were unneeded; they are attributes that do not provide value to the project. Those columns were eliminated or dropped as a result. A feature with too many rows with the same value is an example of data that may not be useful to a study. The categorical factor that indicated whether or not the individual smoked in our scenario was eliminated because the whole column had the value ‘No’.

3.9.3 Data Transformation

Data transformation involves several processes; modifying data and turning them into a proper format is one of the ways. An example is changing data from a string format to an integer format. Another form of data transformation is feature selection. This procedure aids in determining which attributes are most useful for data analysis and, as a result, machine learning model training (Pandey et al., 2020). Another technique used in data transformation is to change the row labels, this is accomplished by indexing the data. Data discretization is another type of data transformation; it is a method of grouping a specific category of data into distinct buckets (Tsai& Chen, 2019).

3.10 Model Evaluation

Evaluation is the process that confirms if the prototype works or if it needs some refinement, Ryan (2013) by allowing the designer and client to assess the viability of the system. It allows the designer to make improvements on potential faults of the system. This process allows concept designs to be evaluated fully. Holdout evaluation was used to test the performance of the model. Holdout approach is done by diving the data into the training set, validation set and test set. This is done at the beginning, before inputting data to train the model. The test set, which is the unseen data was used in this section. It is meant to be a substitute for the data in the real world. This shows whether the trained model will perform as expected. That is, testing the model's performance, whether it is able to predict the individuals BP according to their future calendar events.

Evaluating the performance of the machine learning algorithm involves testing the performance using some model evaluation metrics. Mean Square Error (MSE) and R squared (R2) are the metrics that were used for evaluation of the regression model. MSE is the average of the squared difference between the actual value and the predicted value (Seif, 2019). It is the standard way of calculating the cost function for gradient boosting regressor. The values here lie between 0 to infinity, the smaller the MSE the better the model.

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2$$

MSE - Represents the mean squared error

n - Represents the number of data points.

y - Represents the observed values.

$\hat{y}$ - Represents the predicted values.

R^2 score is one of the metrics used to evaluate the performance of a regression machine learning model. It measures the amount of variance in prediction compared to the actual values. The values of R^2 fall between 1 and 0. 1 means that the model is perfect, while 0 means that the model will perform poorly on unseen data (Kharwal, 2021).

$$\square 2 = 1 - \frac{\square\square\square}{\square\square\square}$$

SSr- Represents the squared sum error of regression line.

SSm - Represents squared sum error of mean line.

3.12 Ethical Considerations

A researcher should maintain moral standards while conducting their study. Therefore, the researcher ensured that the participants of the study had prior adequate information from which they could choose to participate or not. A letter from the institute of Post Graduate and Research Studies of Kabarak University as well as a permit from (National Commission for Science Technology and Innovation) NACOSTI vetted**,** to make sure that the ethical considerations were ensured. The privacy and confidentiality of the respondents' information were protected by strict ethical standards. All research participant data will be handled with strict privacy protocols, including encryption, limited access, and anonymization, ensuring that personal information remains confidential and securely protected throughout the research process. The researcher also assured the respondents that the research is only for academic purposes.

CHAPTER FOUR

DATA ANALYSIS, PRESENTATION AND DISCUSSION

4.1 Introduction

This chapter present the findings of the data collected, the interpretations of the data, and the discussion as per the research objectives presented in section 1.5.

4.2 The Development of a Model for the Regular Collection of Blood Pressure Readings and Activity Data

This section presents the results for objective one of the study; that is to develop a model for regular collection of blood pressure (BP) readings and activity data. Data was collected in Uasin Gishu County in all of its six constituencies. There were 45 participants who accepted to participate in the study. Data collection via the application was successful but not without challenges. The application was able to capture the data needed for the study, which were mainly the BP data, the heartrate (HR), mood and activities. The data collected was safely stored in the cloud for ease of retrieval.

Rapid Application Development (RAD) was used to develop the application called Smarthealth which was used to collect data from the participants. The feedback received from the participants in each iteration of the application informed the development process. The application allowed for data retrieval from both the smartwatch and the participant.

4.2.1 The Smarthealth Application

This section shows the overall look of the application, why the various parts of this application were developed and the results of this exercise.

a) Demographic Information

This is the information that the participants provided after the installation of the application. They provided their name, their date of birth (D.O.B), email address, gender, height and weight for the purposes of BMI, the family history of hypertension, and finally, if they have contracted some of the conditions that may affect BP. These conditions are kidney disease, diabetes and lupus to name a few.

b) Daily Information

There are some questions which the individuals also answered at the end of every day, these are: The amount of sleep they had the previous night, if they have exercised that particular day, if they have taken any medication (and the number of times on that particular day), if they have had alcohol and whether they have smoked and if so, how may cigarettes.

c) Information Collected in 2 Hour Intervals

Every 2 hours, the application prompted the participants to be able to enter their mood and activity data. The mood section asked them how they felt at that particular moment, while the activity data section asked them what they had been doing in the past 2 hours. This information was answered during the day in 2-hour intervals, or when the participants were awake and not when they slept for long hours.

d) BP and HR data

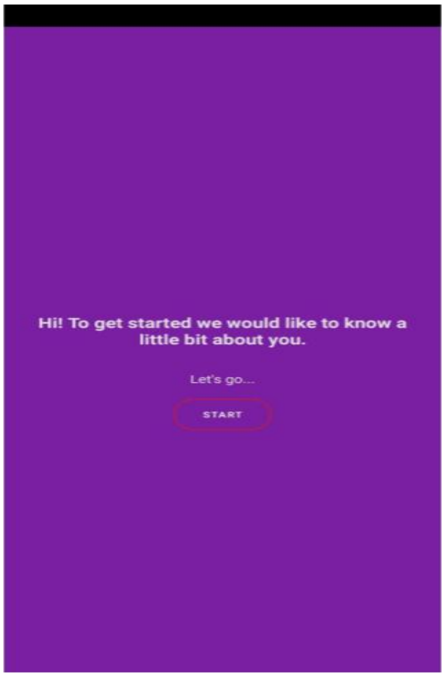
The BP and HR data were measured after every 30 minutes, both day and night. This was possible because the application was designed to prompt the smartwatch to take the measurements automatically. Automation was better because it saved time for the participants and this way, more measurements were taken because manual measurements would be forgotten as seen earlier in the pilot study, where it was partially automatic. Automation also made it more convenient for users because they didn't have to think about taking measurements manually and finally giving up on taking measurements because of frustration.

e) The Smarthealth Application

These are pages of the Smarthealth application. Each screenshot has a description of the page below it.

Figure 19

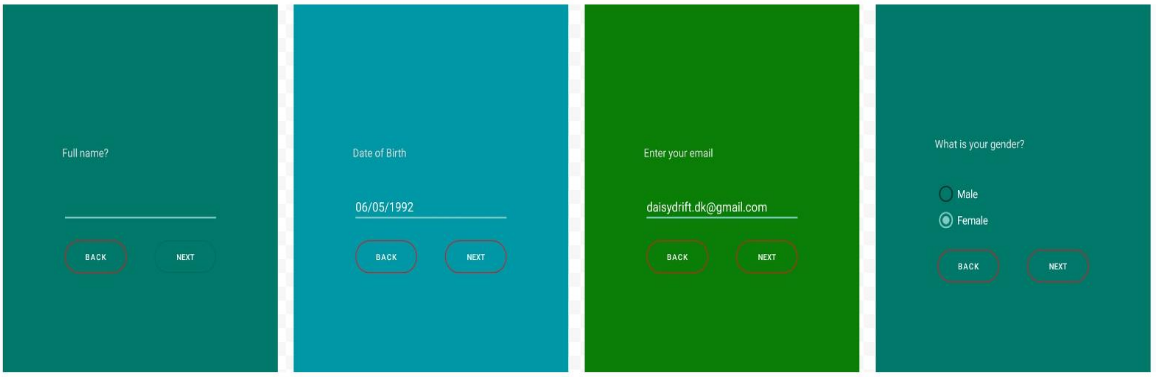
Introduction page of the Smarthealth application



Once the application is installed, this is the first page the participants sees. This page lets them know that they will be providing some information in the application.

Figure 20

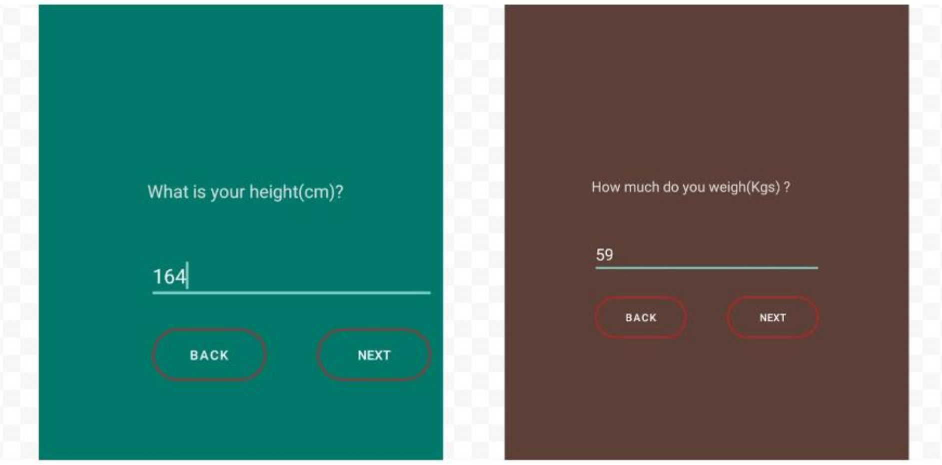
The pages that prompt for individuals' demographic information



The pages that follow are the name page, the date of birth page, the email page and the gender page.

Figure 21

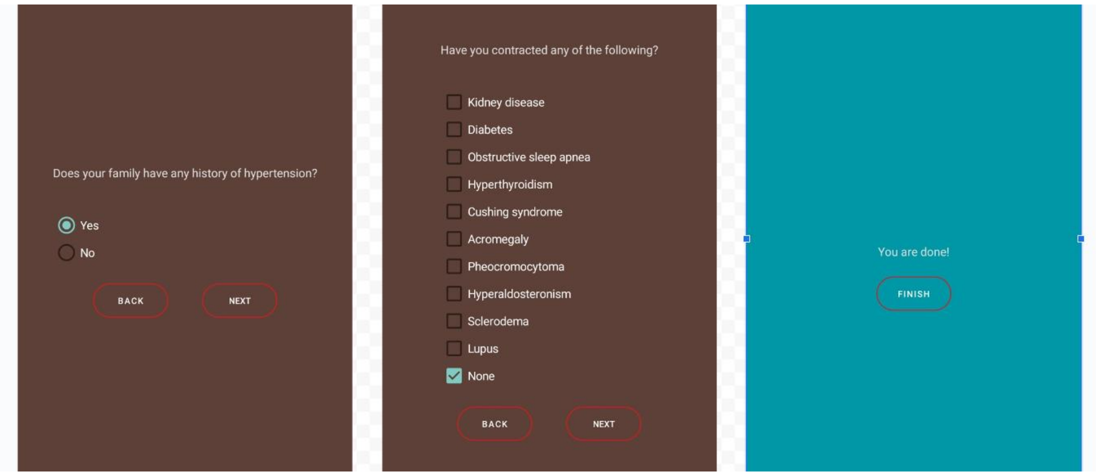
The height and weight pages



The next two pages are of the participants' height and weight. The two are used to calculate the BMI of the individual.

Figure 22

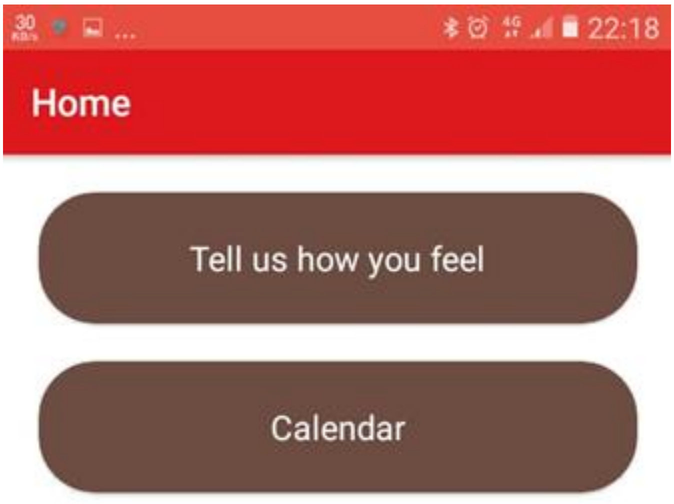
The final pages of the introductory part of the application



The page that follows the weight page asks the individual whether they've had a history of hypertension in their family. The next page asks the participant whether they have certain medical conditions. The final page informs the participant that they are done with the on boarding process.

Figure 23

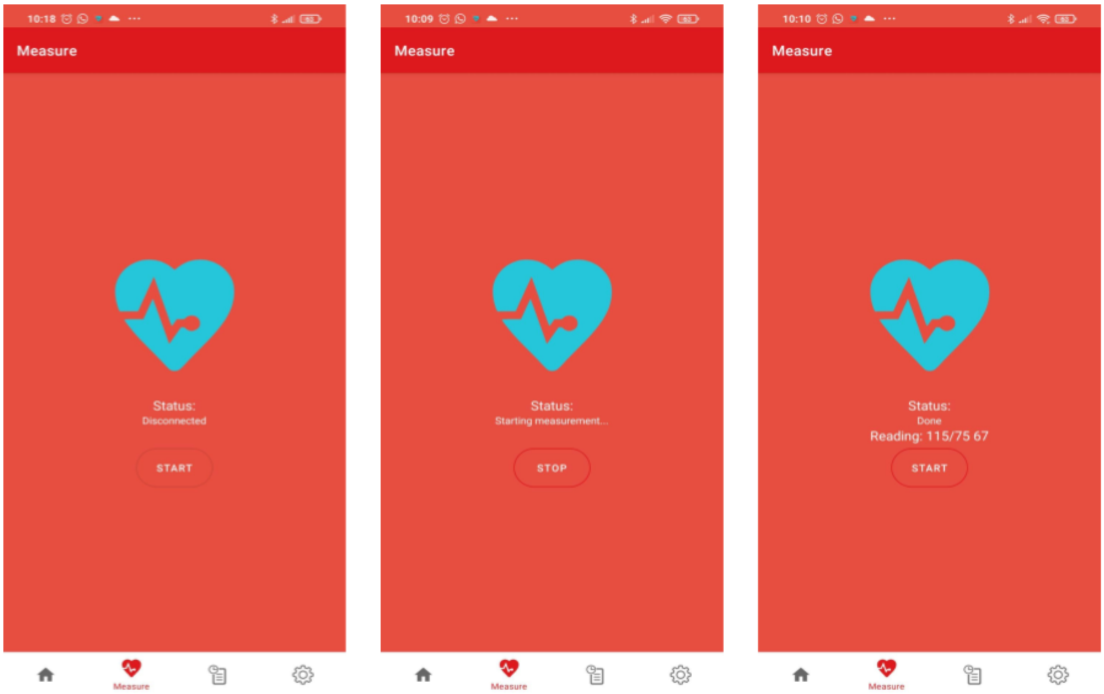
Home Page



This is the home page of the application; this opens up once the user is done entering the information needed. The top section has two options, one for the user to manually enter their mood, the other is for the user to enter a calendar event. The bottom has four options, Home for home page which is the current page shown in figure 23, the next one is the measurement page, the one that follows is the history page, and finally the settings page.

Figure 24

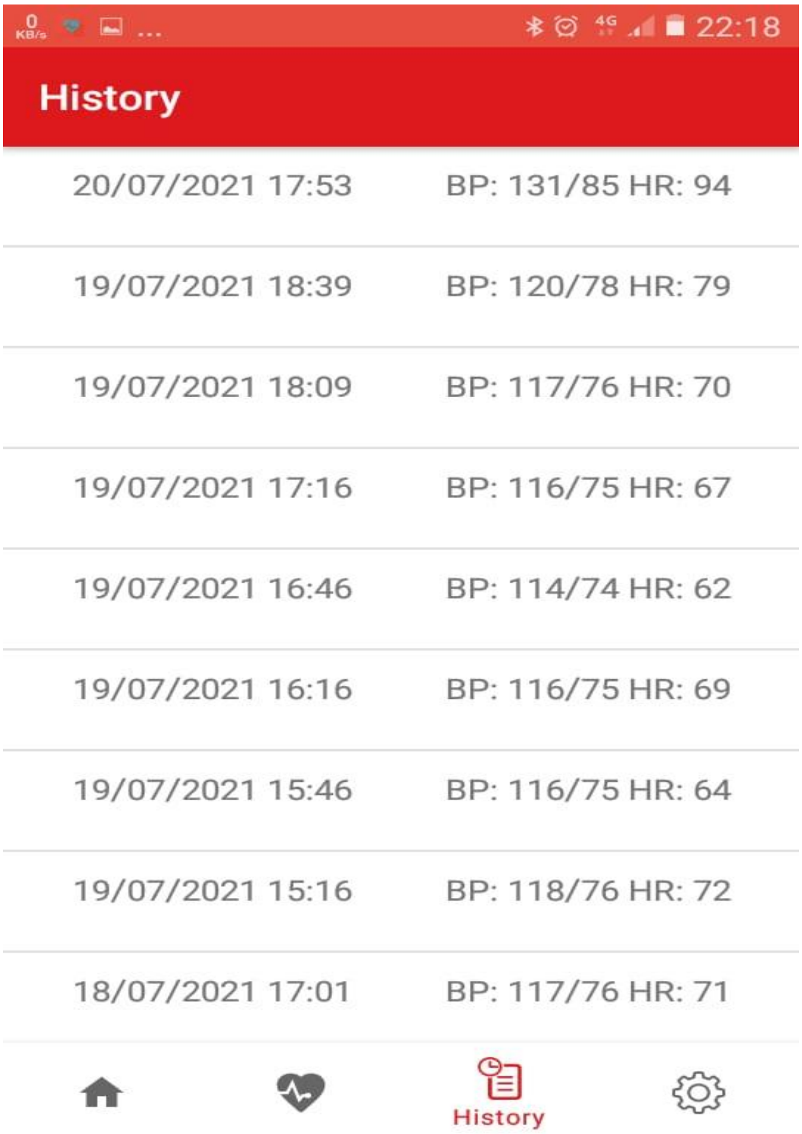
The Measurement Page



This is the measurement page; a user can manually measure their BP by pressing ‘START’. When a user’s watch is disconnected, the status will indicate like the first image in figure 24above. When the device is connected, it indicates “Starting measurement”, and when the measurement is done, it indicates the SBP, DBP and HR respectively as can be seen in the third image in the figure 24above.

Figure 25

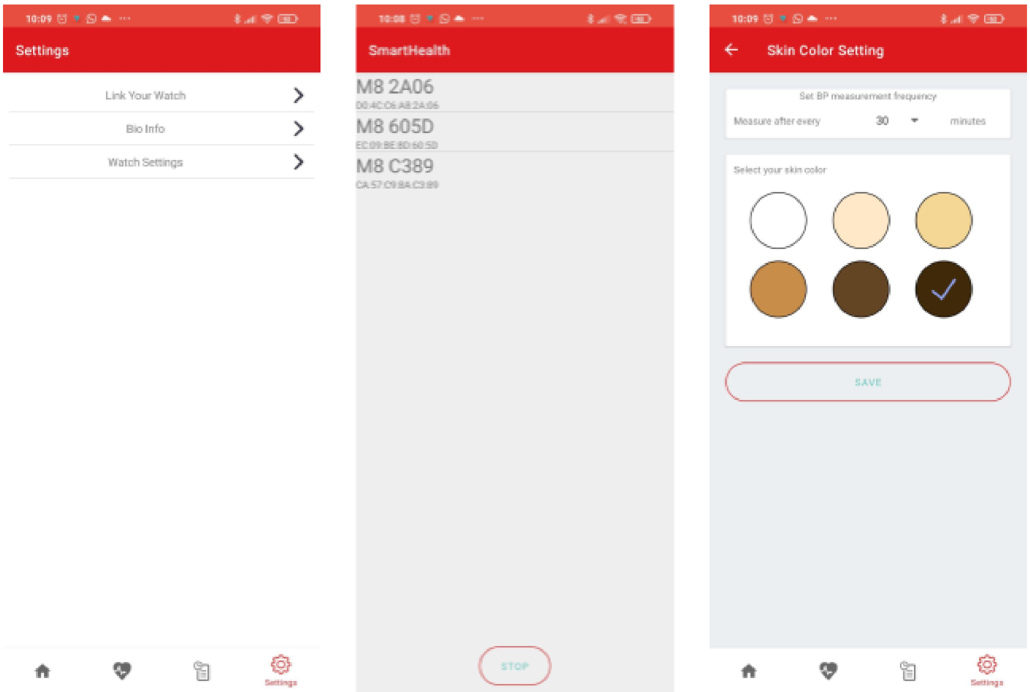
The history page



This is the history page; it displays the participant's BP and HR together with the date and time the measurements were taken. Users use this page to monitor their BP over time.

Figure 26

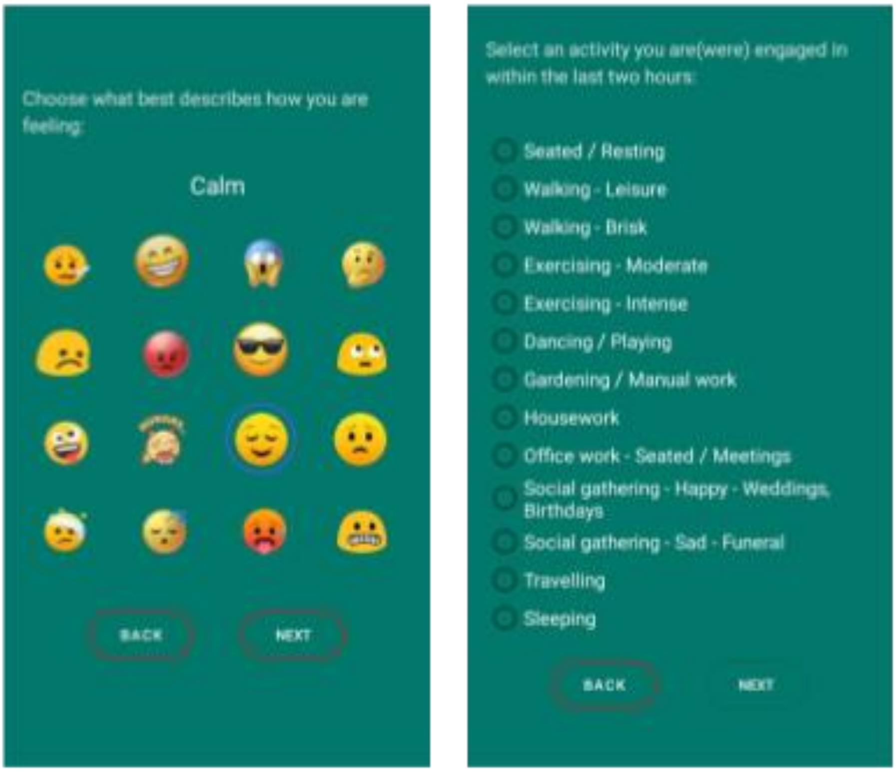
The Settings Page



The first image in Figure 26 above shows the settings page; there are three options displayed, the first which is “Link Your Watch “which provides an option to link the watch, it can be seen in the second image. When the user presses “Scan” the application scans for the smartwatches that are nearby via Bluetooth and displays their IDs, the user can then link the smartwatch using its own unique ID. The user can view his or her information when they click on “Bio Info”; this shows the summary of their demographics. The third option which is “Watch Settings” has options for choosing the skin colour and time interval for measuring the BP and HR. The time interval used in this study was 30 minutes.

Figure 27

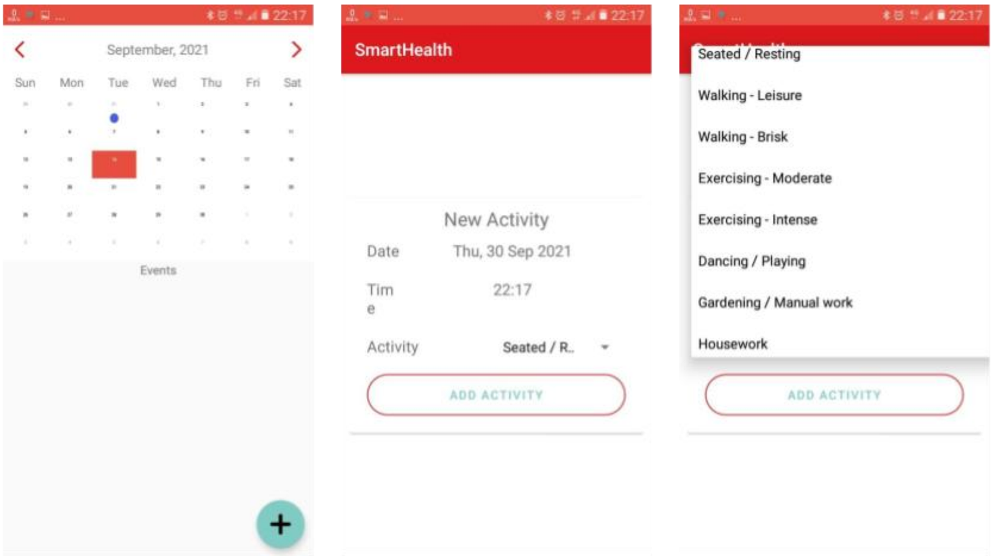
The mood and activity pages



There are two pages that pop up consecutively after every 2 hours on the user's smartphone; the first one, which is the first image in figure 27ask the participant how they feel, their mood. The second one, which is shown by the image on the left allows for the participant to choose the activities they may have been involved in in the past 2 hours, from the 13 options displayed.

Figure 28

Calendar Events Page



When a user clicks on the "Calendar" option, located in figure 23, a page displaying a calendar is presented, as depicted in the first image shown in figure 28. The participant then chooses a date, then clicks on the plus sign which is below the page. The page that follows is used to select a date and time for a future calendar event. Once the user is done choosing the time, they press on “Add Activity”. The page that follows allows the participant to select the activity they may be involved in at a particular future day and time. The activities displayed in this page are similar to the activities displayed in the second image in Figure 27.

4.2.2 To Develop a Model for Regular Collection of Blood Pressure Readings and Activity Data

The first objective was to develop a model for regular collection of blood pressure readings and activity data. This section discusses the results of the data collection process and the findings of the data collected.

Inputs from the participants: Advantages

i. The Bluetooth connectivity worked well for 2/3 of the participants. This made automation easier thus collecting more data.

ii. Participants liked the fact that the watch reduced the need for them to go to the hospital to take their BP readings. They liked the fact that they could take BP anytime, anywhere.

iii. Some participants said that they didn't know their BP because they rarely got checked, so it proved to be very helpful to them because they now had knowledge of their BP status.

iv. For some of the participants, when they noticed that the pressure was higher than usual, due to anger for example, they said that they would try to calm down. The application helped them keep their emotions in check.

v. Some participants found that the application was a good motivator for doing exercise. Whenever they noticed a rise in BP, they were motivated to try and reduce it by exercising.

vi. For most participants, the watch battery lasted about 3 days, which they considered good battery life.

vii. Some participants said that the application helped them stay alert, live more consciously, they were more cautious of what they let themselves feel, for example.

viii. Almost all the participants wanted to either remain with the watches, or buy them, they liked its functionality. Some referred friends who then wanted to take part in the study for the benefit of their health.

Inputs from the participants: Disadvantages

i. The connection for some of the participants, about 1/3, was poor, especially after one week of use. So they resorted to measuring the BP and HR manually

   This resulted to having inconsistent data, especially when it comes to time. The smartwatch, if well connected, records BP reading and heartrate every 30minutes. Unfortunately, when disconnected it becomes hard for the participants to keep recording the data after every 30minutes because of a busy schedule or forgetfulness. This becomes impossible at night because most participants are asleep.

ii. Some participants found it stressful to sleep while wearing the watch

   Not everyone usually wears an accessory around their wrist, those that are not used to wearing accessories found it more difficult to sleep with the smartwatch at night. This also led to missing of data for some of the participants.

iii. Some participants, especially those who do not wear accessories on their wrists, found it vexing

   As mentioned, in (ii), not everyone likes to wear wrist accessories. A few participants did have complains at the end of data collection about this. This however, did not prevent them from recording data during the day.

iv. The application drained the smartphone battery fast

   The Bluetooth connectivity between the smartphone and the smartwatch drained the smartphone battery fast. During the 2 weeks of data collection, some individuals had to charge their phones more often then they usually did. Due to this, the participants may have forgotten to record the data that was collected every 2hours because they didn't have their phone with them.

v. The smartwatch vibrates at times, especially after detecting the individual's temperature is quite high, that tends to surprise the participants

Some of the participants complained that watch vibrations due to high temperature did surprise them, especially if the smartwatch vibrated during the night.

vi. Some participants were concerned about the green light (PhotoPlethysmoGraphy) emitted while taking some readings like BP and HR

Photoplethysmography (PPG) is an optical measurement method that is often used for heart rate monitoring purposes. Castaneda *et al*. (2018) defined it as a "non-invasive technology that uses a light source and a photodetector at the surface of skin to measure the volumetric variations of blood circulation" (p. 195). So far, there is no research that indicates any harm brought about by this technology. Also, after the data collection process, no participant brought up any issue concerning the PPG light affecting them in any way.

Conclusion of the findings for objective one

i. Turning off the location of the phone made it more difficult for the application to automatically prompt the smartwatch to take the readings

Some of the participants were concerned about turning on location services on their phones. They feared being tracked. The application does not track individuals; the sole purpose of the GPS was to make it easier to connect with the smartwatch.

ii. Some participants did not communicate when there was a problem with the watch or the Smarthealth application, this led to inconsistent data

Due to lack of communication, whenever there was a problem, an example would be charging the smartwatch, data for that particular day wouldn't be collected. This led to incomplete data.

4.3 To Investigate of a Suitable Machine Learning Algorithm for the Prediction of Future Blood Pressure Readings of an Individual Using Their Past Blood Pressure Readings and Activity Data

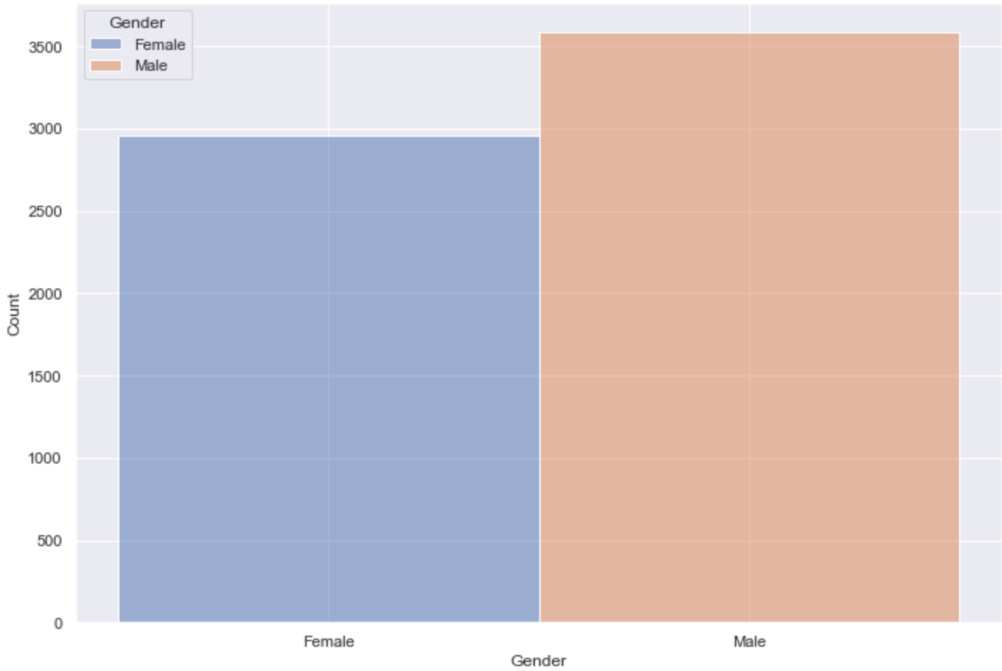
In this section, an exploration will be conducted on both the descriptive statistics and the inferential statistics of the data collected using the smartphone application and the smartwatch. The process of investigation of a suitable model will then be demonstrated.

4.3.1 Descriptive Statistics

MySQL and Seaborn, a statistical plotting library, were used to aid in describing some characteristics of the participants in the study.

Figure 29

The graph represents the gender count of this group of interest.



Men are more likely to develop high BP compared to women, as stated in section 3.6 of chapter 3. The data collected therefore consisted of more males than females as can be seen in Figure 29. The number of male participants were 28, and 17 were female.

Figure 30

The graph represents the age count of this group of interest.

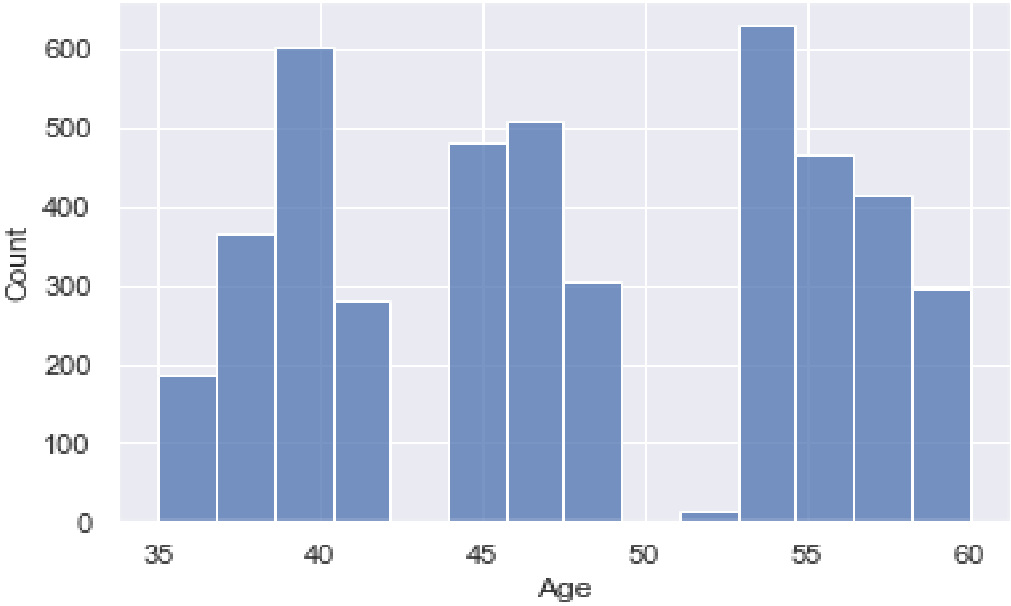
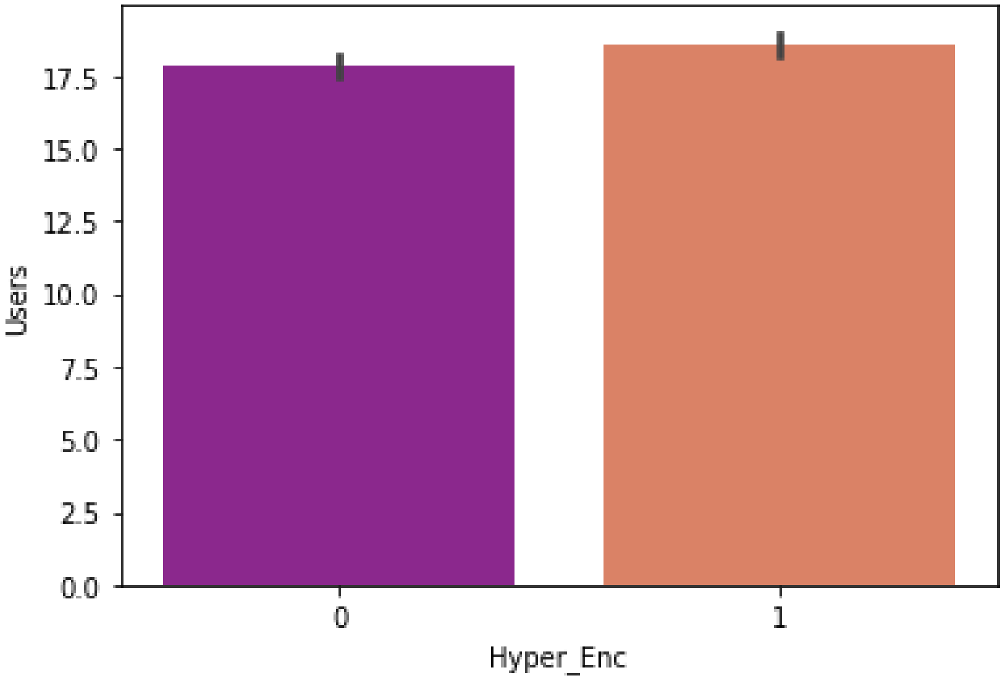


Figure 30 shows a histogram used to display the age of the participants. The age of the participants was between 35-60 years, the average age being 45. This is because study shows that hypertension tends to be developed at the age of 35 and above.

Figure 31

The graph represents the hypertension history of this group of interest.



In Figure 31, the data has been encoded, those that do not have a history of hypertension are represented by is 0 and those that do have a history of hypertension are represented by is 1. This graph shows that nearly half of the population have parents or relatives who have or had hypertension.

Figure 32

The graph represents the BMI of this group of interest.

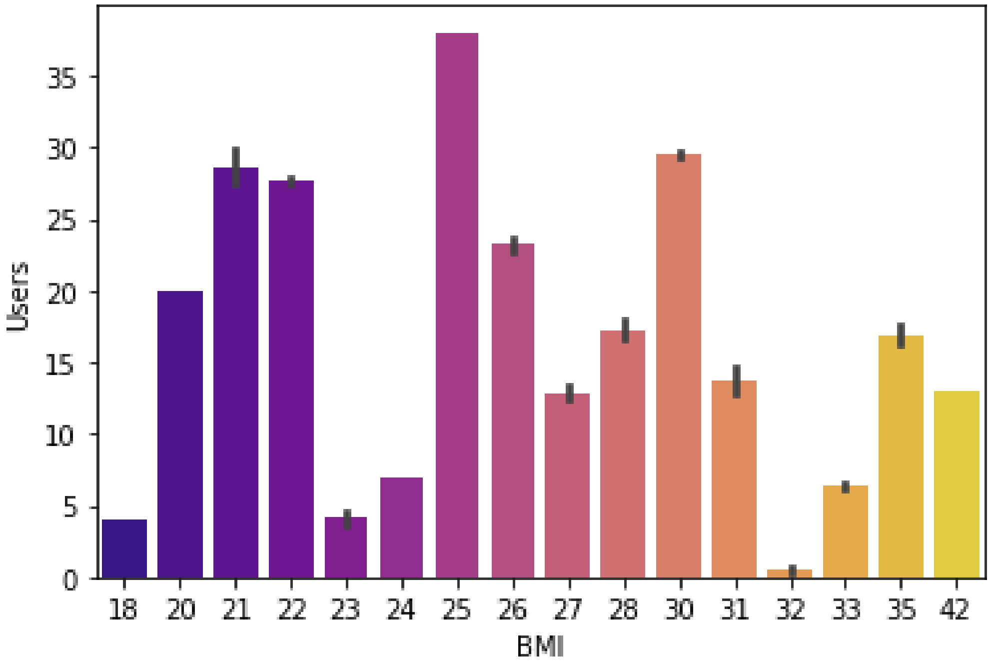


Figure 32shows users and their BMI. We can see that there is a wide range of BMI categories. This is the BMI interpretation table according to the CDC.

Table 4

BMI interpretation for adults

BMI	Weight Status
Below 18.5	Underweight
18.5 – 24.9	Healthy Weight
25.0 – 29.9	Overweight
30.0 and above	Obesity

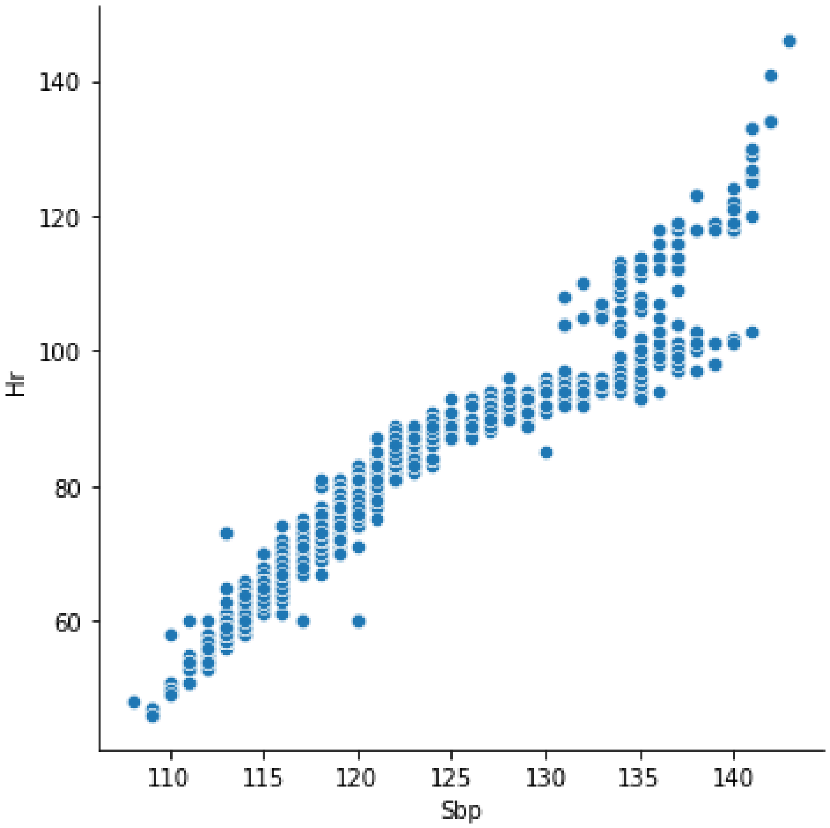
Source: Zierle-Ghosh & Jan, (2023)

4.3.2 Inferential Statistics

MySQL and Seaborn, were used to aid in describing this particular group of interest.

Figure 33

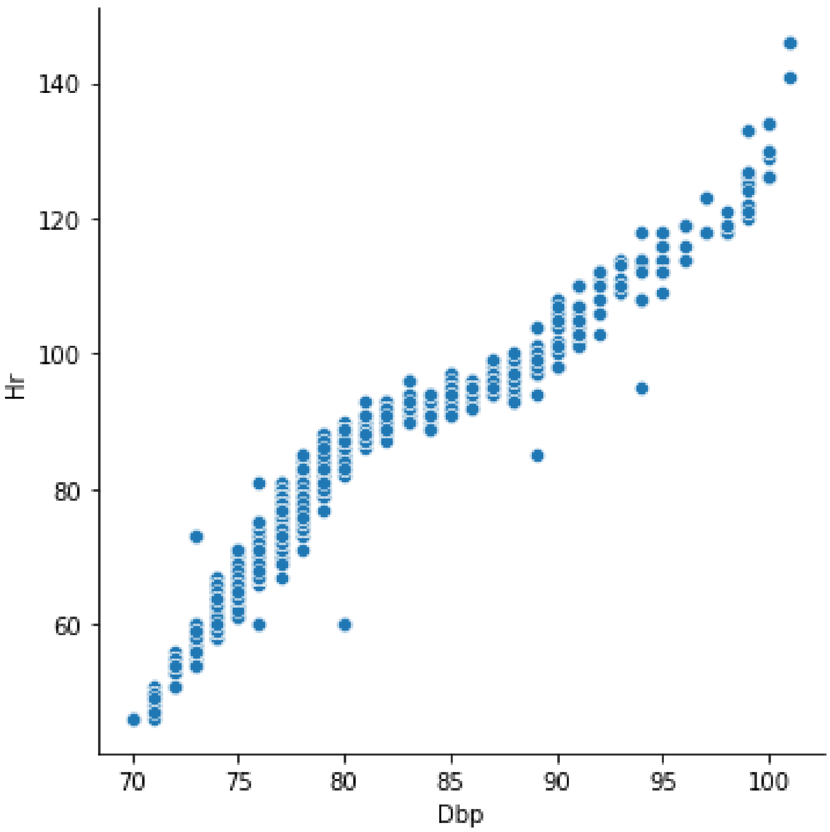
The graph represents HR against SBP



This relational plot, Seaborn's Relplot() in particular, shows the relationship between two variables, heartrate (HR) and systolic blood pressure (SBP). From the relational plot, it is observable that as HR increases, the SBP also increases. This therefore shows that there is a high correlation between the HR and the SBP.

Figure 34

The graph represents HR against DBP



This relational plot also shows the relationship between two variables, HR and diastolic blood pressure (DBP). As we can see in the relational plot, as HR increases, the DBP also increases. This also shows that there is a high correlation between the HR and the DBP.

Figure 35

The graph represents the DBP and SBP against time

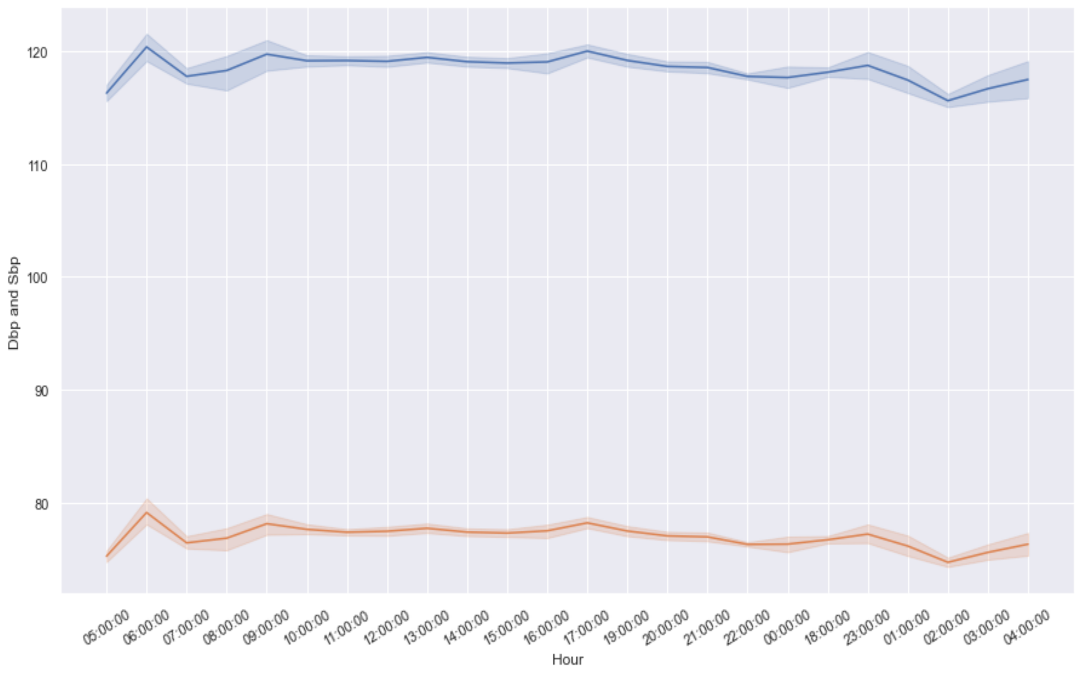


Figure 35 shows that there is a high correlation between the DBP and SBP. It also shows that both the DBP and SBP are highest in the morning hours, at around 6:00am, and is lowest at night at around 2:00am, when most individuals are asleep.

Figure 36

The graph represents HR against time

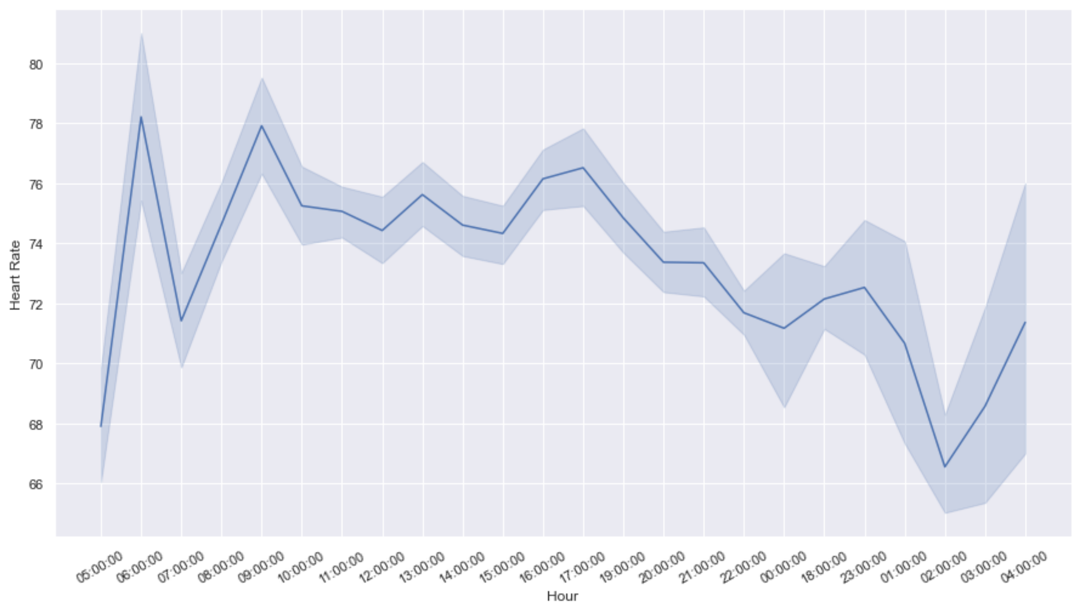
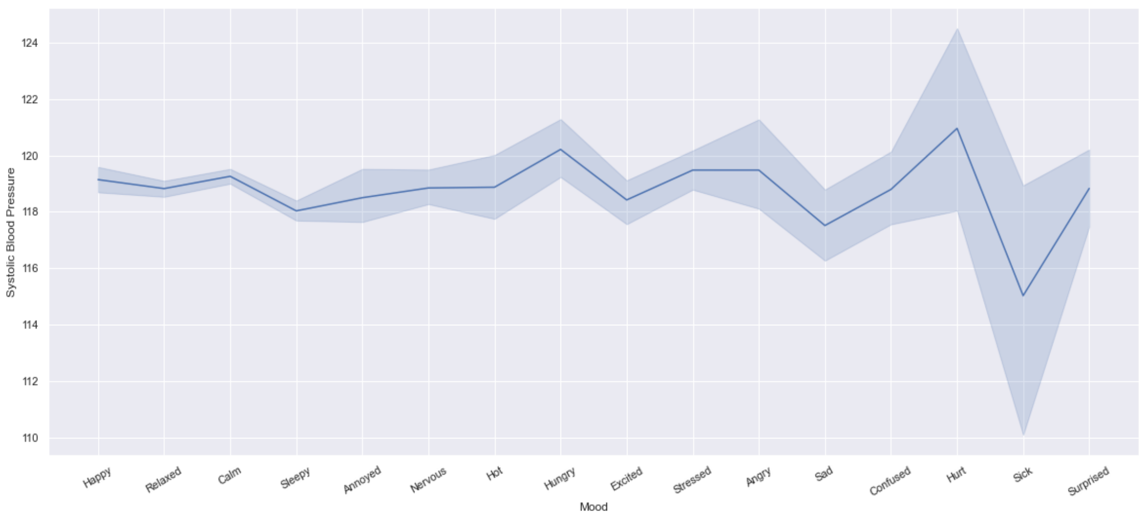


Figure 36 also shows that HR raises in the morning hours and is low at night, when most individuals are asleep. It also reveals that from around 5:00pm, the HR gradually reduces up until around 2:00am, when it is lowest, then it raises. Figures 35 and 36 show that there is a high correlation between the three variables, the HR, the SBP and the DBP.

Figure 37

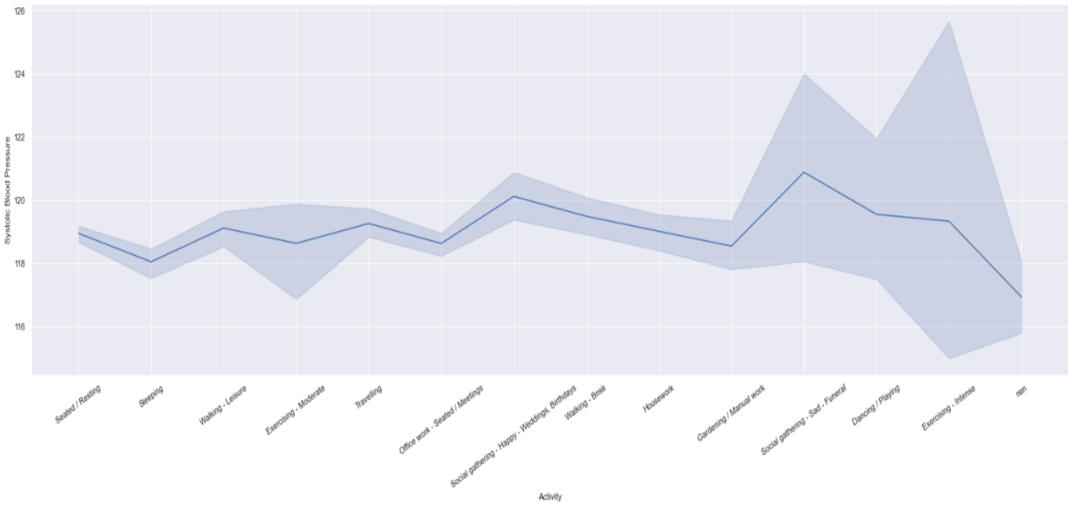
The graph represents the SBP against the mood.



This line plot in Figure 37 shows that the mood affects the SBP. An example would be when an individual is sleepy, the SBP goes low up to about 118, and when an individual is hurt, the SBP spikes up to 124.

Figure 38

The graph represents the SBP against activity.



This line plot in figure 38also shows that the activity affects the SBP. An example would be when an individual is sleeping, the SBP goes low up to about 118, and when an individual is in a social gathering that is relatively happy, the SBP spikes up to 121.

To further analyse the data collected, multicollinearity was carried out. Multicollinearity occurs when several variables are significantly correlated not only with the dependent variable but also with each other in a multiple linear regression analysis. The variance inflation factor (VIF) test was used to assess multicollinearity. VIF determines the strength of the correlation between independent variables (Shrestha, 2020). The formula for VIF is:

$$\square\square\square = \frac{1}{1 - \square_{\square}^{2}}$$

To calculate the VIFs, all independent variables are converted into dependent variables. Each model produces an R-squared number that represents the proportion of the individual's variation explained by the set. As a result, higher R-squared values indicate greater multicollinearity. In VIF calculations, these R-squared values are used (Shrestha, 2020).

The VIF test found significant correlations between the SBP, DBP, heart rate, BMI, hour, and age. Multicollinearity is not a problem in this case because the BP of an individual is predicted using all the features rather than just one.

4.3.3 The Data with Regard to Machine Learning Models

There are several types of data in machine learning. The three most common data types are; cross-sectional data, time series data and panel data. Cross-sectional data is one that is set at a fixed point in time. It is characterized by individual units like countries, people or different houses (Lawson et al., 2018). The dataset usually contains multiple variables. An example of cross-sectional data is obtaining end of year scores for students, other factors can be included like how often they attend classes.

Time series data is a sequence of measurements of the same variable collected over time. Time is used to identify each observation (Tavenard et al., 2020). An example is taking the oxygen level measurements of a single patient once a day, over a period of 6 months. Panel data usually contains different cross sections across time. It is a combination of cross-sectional data and time series data. Time series is unidimensional while panel data is multidimensional. It focuses on multiple individuals at multiple time periods (Bliese et al., 2020). The time periods can be measured in years, months, weeks, hours, minutes or seconds. An example of panel data is individuals' yearly income over time across different counties in Kenya. This data could factor in age, sex, education level and other variables.

Table 5

An example of panel data showing income per year across different individuals from different counties in Kenya

Person	Year	Income (KES)	Age	Sex	Education Level	County
1	2019	240000	31	M	Diploma	Bungoma
1	2020	240000	31	M	Diploma	Bungoma
1	2021	240000	31	M	Diploma	Bungoma
2	2019	360000	23	F	High School	Kilifi
2	2020	360000	23	F	High School	Kilifi
2	2021	360000	23	F	High School	Kilifi
3	2019	6000000	57	M	Masters	Isiolo
3	2020	6000000	57	M	Masters	Isiolo
3	2021	6000000	57	M	Masters	Isiolo

Source: Researcher (2023)

The data collected in this study comprises of panel data because it measures multiple variables of multiple individuals across multiple time periods. However, there is complications with the data collected, the time is not in even intervals, therefore making the data inconsistent. The time periods as mentioned in section 4.2.2, are not continuous for all the participants. This forms an unbalanced panel data which is difficult to balance out because of the extensiveness of some incomplete data. These findings lead the researcher to go ahead with the panel data but using time as a numeric format instead of date-time format and added the time aspect as one of the independent variables.

4.3.4 The Investigation of a Suitable Machine Learning Model

The researcher used Anaconda, which is a Python distribution platform for scientific computing, and Jupyter Notebook which is a web application that allows for creation of code, for coding purposes. After the data was cleaned and analysed, data pre-processing was done so that the models would take in the given data. Most machine learning models cannot do well with unprocessed data. Feature selection was then done in order to reduce dimensionality. This process involves isolating data that is consistent, non-redundant and relevant. Pandas Data Frame Correlation was used; this method finds the correlation of columns in a data frame. Highly correlated columns were then used as data for the machine learning model. This in turn gives the model a better chance of performing well.

Using machine learning libraries like Scikit-learn, the researcher was able to test several models in order to find one that works well with the data collected. The models tested were Lasso, Linear Regression, and Elastic Net to name a few. Label encoding was used to convert categorical data to numerical format. Label encoding was used in most columns due to the large number of variables in those columns. The data was then split into the dependent variable and the independent variables. In this case, there are two dependent variables, the SBP and the DBP of an individual. For the purposes of training and testing, the data was split into training data which took 70% of the data and testing data which took the remaining 30%. A machine learning pipeline was then used to automate the procedure it takes to produce a machine learning model. A pipeline consists of multiple steps from data extraction, the pre-processing of data, to model training and even deployment. In this case, pipelining was mainly used for standardization of data, it was also used to test the possible models and select one that performed the best, and for model evaluation purposes.

4.3.5 The Results of the Investigated Machine Learning Models

MSE and R^2 are the metrics that were used for evaluation of the regression model. These are the results of the six (6) models that were investigated, trained and tested in order to find the best one for the study:

Multiple Linear Regression is a statistical technique for predicting a variable's outcome based on the values of two or more variables. It is an extension of linear regression and is also referred to as multiple regression (Etemadi & Khashei, 2021). The difference between the equation for Linear Regression and the equation for Multiple Regression is that the latter requires the ability to handle many inputs, whereas Linear Regression requires just one. In Scikit-learn there is no alternative library to implement Multiple Linear Regression it only has Linear Regression, which works for both Linear Regression and Multiple Regression like in this particular case. The MSE score was 0.975 and R^2 score was 0.966.

For Linear Regression models, Lasso regression analysis is a shrinkage and variable selection method, it shrinks the regression coefficients towards zero (Ranstam & Cook, 2018). The purpose of lasso regression is to find the subset of predictors that produces the least amount of prediction error for a quantitative response variable. Lasso accomplishes this by imposing a constraint on the model parameters that leads some regression coefficients to decrease toward zero. Lasso had a MSE score of 1.157 and R^2score of 0.957.

Elastic Net is a regression approach that simultaneously does variable selection and regularization. The basic principle underlying the Elastic Net is regularization; it is a series of strategies that can assist minimize overfitting in neural networks, enhancing the accuracy of deep learning models when they are fed new data from the issue domain (Giglio& Brown, 2018). Elastic Net had a MSE score of 0.927 and R^2 score of 0.961.

The K-Nearest Neighbours (KNN) algorithm predicts the values of new data points based on 'feature similarity.' This means that a value is assigned to the new point based on how similar it is to the points in the training set (Triguero, García- Gil, Maillo, Luengo, García, & Herrera, 2019). Scikit-learn's Kneighbors Regressor had a MSE score of 0.773 and R^2score of 0.971.

A decision tree constructs regression or classification models (in this case, a regression model) in the form of a tree structure. It incrementally cuts down a dataset into smaller and smaller sections while also developing an associated decision tree (Fiskin, Cakir, &Sevgili, 2021). A tree with decision nodes and leaf nodes is the end result. The Decision Tree Regressor had a MSE score of 0.219 and R^2score of 0.992.

Gradient boosting is a type of ensemble method in which generates numerous weak models and combines them to improve overall performance, it is a type of machine learning boosting. It is based on the assumption that when the best potential next model is coupled with prior models, the overall prediction error is minimized (Cai, Xu, Zhu, Hu, & Li, 2020). Gradient Boosting Regressor had a MSE score of 0.182 and R^2 score of 0.992.

Out of all the identified models, Decision Tree Regressor and Gradient Boosting Regression performed the best with an accuracy of 99%. The MSE was 0.219 and 0.182respectively and the R^2 score of both models was 0.99. Gradient Boosting Regression was used for the regression model because of its lower MSE compared to the Decision Tree Regressor.

Table 6

Results of Each Identified Model

Model Name	MSE score	R^2 score
Multiple Linear Regression	0.975	0.966
Lasso Regression	1.157	0.957
Elastic Net	0.927	0.961
K-Nearest Neighbours	0.773	0.971
Decision Tree Regressor	0.219	0.992
Gradient Boosting Regressor	0.182	0.992

Note. The table shows the MSE and R^2 scores of each model.

4.4 To Train the Suitable Machine Learning Algorithm to Predict Future Blood Pressure Readings of an Individual Using Their Past Blood Pressure Readings and Activity Data

After the investigation of the models, Gradient Boosting Regressor (GBR) was identified as the best model out of the six that were investigated. A predictive model was then created using the GBR algorithm.

The following steps were followed:

Step 1: Essential machine learning libraries, such as Pandas, Scipy, and Sci-kit learn, were imported.

Step 2: Subsequently, the data was imported.

Step 3: The data underwent a comprehensive cleaning process.

Step 4: Feature engineering was performed at this stage.

Step 5: Pertinent attributes were selected using the Pearson Correlation.

Step 6: The data was then divided into training and test sets.

Step 7: The GBR model was defined and subsequently fitted.

Step 8: The model was trained using the prepared data.

Step 9: Finally, the trained model was then tested against the test data.

The following figures show the result of the prediction model when compared to the test data.

Figure 35

This image shows the SBP prediction, the test data and their difference.

Out[34]:

	Prediction	Test Data	Difference
0	118.186517	118	-0.186517
1	114.195993	114	-0.195993
2	115.098294	115	-0.098294
3	119.853309	120	0.146691
4	116.238001	116	-0.238001
5	124.926891	125	0.073109
6	114.347829	115	0.652171
7	110.343997	110	-0.343997
8	137.079042	137	-0.079042
9	119.895421	119	-0.895421

This is the prediction of SBP. As can be seen in figure 35, the difference between the predicted data and the test data is very small.

Figure 36

A graph showing SBP predicted values against the actual values

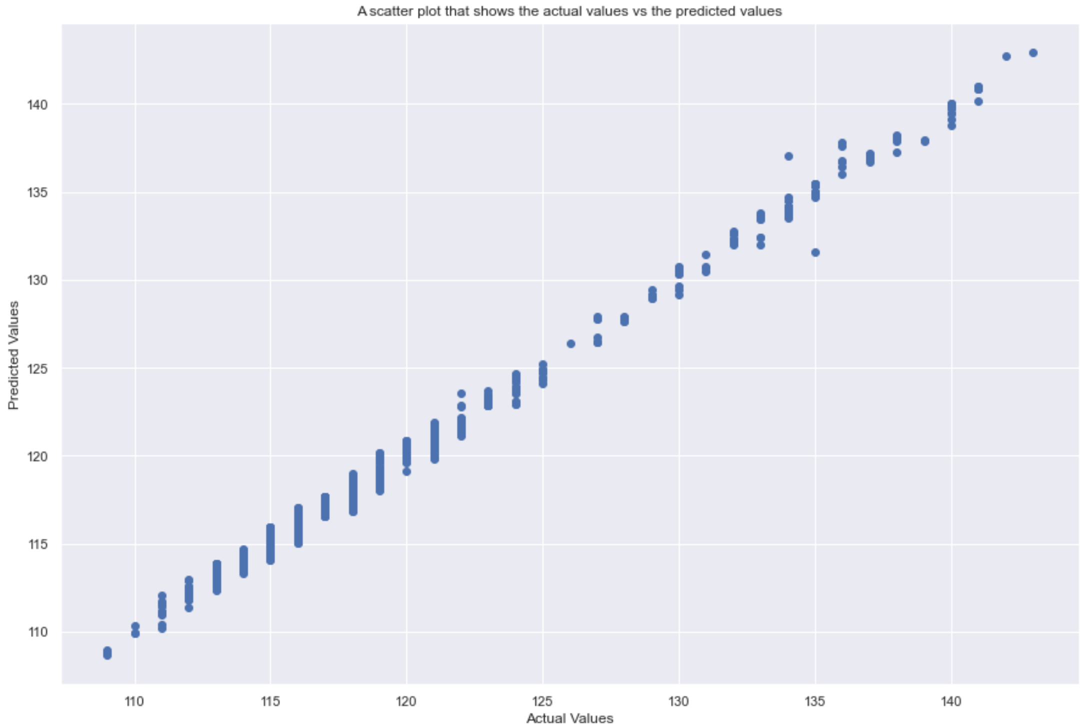


Figure 37

This image shows the DBP prediction, the test data and their difference

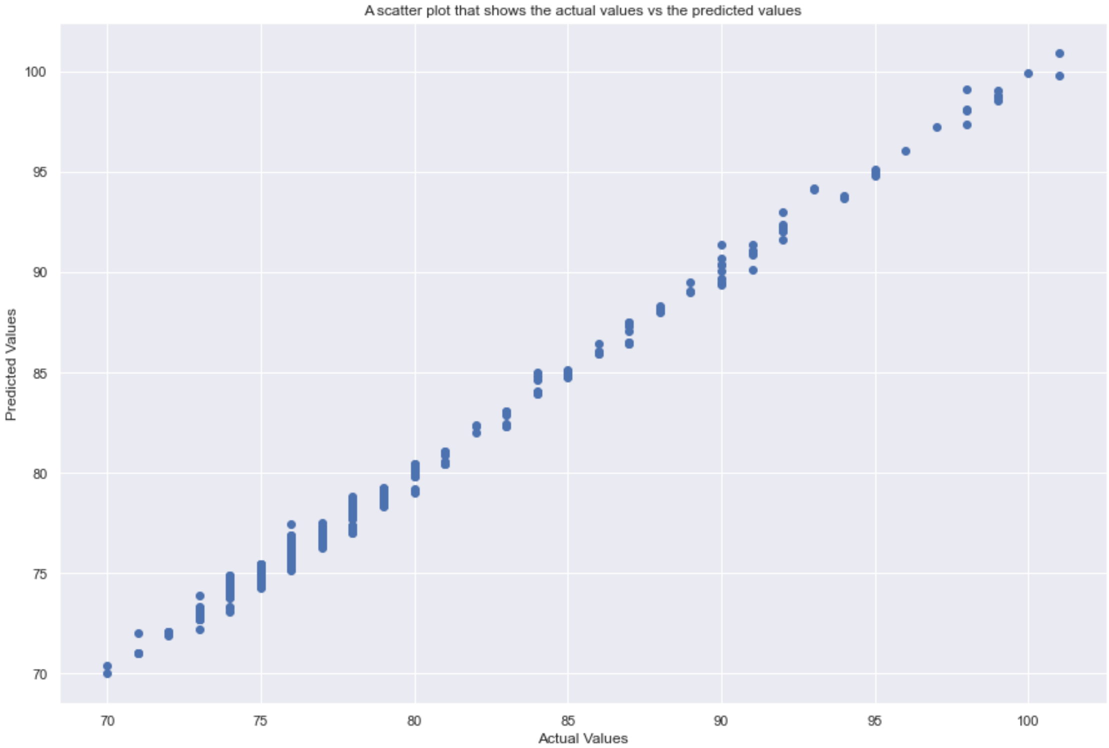
Out[20]:

	Prediction	Test Data	Difference
0	76.931912	77	0.068088
1	73.975209	74	0.024791
2	74.489936	75	0.510064
3	77.975605	78	0.024395
4	75.298633	75	-0.298633
5	81.035571	81	-0.035571
6	74.620231	74	-0.620231
7	71.006671	71	-0.006671
8	88.999732	89	0.000268
9	77.247206	78	0.752794

This is the prediction of DBP. As can be seen in figure 37, the difference between the predicted data and the test data is very small.

Figure 38

A graph showing DBP predicted values against the actual values.



The GBR model did well in predicting the future blood pressure, which entails the SBP and the DBP of an individual.

4.5 To Validate the Performance of the Machine Learning Algorithm

Validation of a model is the process of evaluating a model on test data. This process provides the general ability of the trained model. It is used to confirm that the model can predict an outcome given certain conditions. Mean square error and R^2 have been used to validate this predictive model.

It is difficult to know the best values for hyper-parameters, that is why it is important to tune the hyper-parameters several times in order to achieve a model that is accurate. Grid

Search CV which comes in Scikit-learn's model selection package, has been used to automate the process of hyper-parameter tuning because doing so manually is resource consuming. How Grid Search CV works is that it tries all the given combinations passed and evaluates the model for every combination giving one that performs the best. One of the arguments passed into the GridSearch CV function is "scoring", this is the evaluation metric that a researcher would want to use to evaluate a model. In this case, the mean_squared_error was used. The lower the mean square error, the better. A model that has a mean square error of 0.0 is a perfect model. Using the best estimator, the model produced a mean square error of 0.19 which shows that the model is accurate. R^2score was also used as a means of validation, the closer R^2 is to 1.0, the better the model. The model was able to produce an R^2 score of0.99. The results show that the created model is accurate, it can predict both the SBP and the DBP of an individual at a future point in time given the independent variables that were used.

4.5.1 Blood Pressure Predictions

The researcher used the section on the Smart Health application called Calendar Events to predict the BP of some individuals. The Calendar Events had no Mood section, this is because one cannot predict their moods at a future date, while it is somewhat easier to predict one's future activities. Therefore, these predictions were done without using the Mood feature. The results showed accurate readings in predicting the short-term blood pressure for the individuals.

Figure 39

Accurate predictions of the individuals

ACTUAL PREDICTIONS FOR INDIVIDUALS

```
In [25]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_01 = model.predict([[74, 61, 0, 23, 6, 21, 10, 38]])
         print(individualBP_01)

         [114.01360554]

In [26]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_02 = model.predict([[80, 86, 1, 30, 6, 33, 21, 45]])
         print(individualBP_02)

         [123.14378541]

In [27]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_03 = model.predict([[77, 77, 1, 18, 5, 3, 20, 41]])
         print(individualBP_03)

         [118.82677256]

In [28]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_04 = model.predict([[75, 66, 0, 27, 11, 9, 15, 46]])
         print(individualBP_04)

         [115.58930192]

In [29]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_03 = model.predict([[83, 92, 1, 18, 5, 3, 6, 41]])
         print(individualBP_03)

         [127.56852753]
```

The results in figure 39 shows the variables the researcher used to predict an individual's BP. These variables include the diastolic BP, systolic BP, the participant's hypertension history, their BMI, the activity, mood, the hour the BP was taken and their age. The print statement displays the individuals predicted BP which can be shown by the square brackets below each print statement.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter provides the summary, conclusion and recommendations on the creation and use of the Smarthealth application, the investigation and design of the machine learning model and the validation of the model.

5.2 Summary

This section presents the summary of the main findings of objectives one, two, three and four.

5.2.1 To develop a model for regular collection of blood pressure readings and activity data

The development process of the Smarthealth application was an iterative process that took some months to finalize. This is because whenever the application was tested, there was always something new to be added or an error to be corrected. The final version of the application (Version 3.7) that was used in data collection delivered what it was intended to do. The application entailed the homepage, which showed the participant their BP and heartrate. It also had the settings page, which allowed the participant to connect to their smartwatches and update their information. The application also allowed the participant to update their mood, activities and future calendar events.

5.2.2 To investigate a suitable machine learning algorithm for the prediction of future blood pressure readings of an individual using their past blood pressure readings and activity data.

This process involved using the data collected using the Smarthealth application to investigate the best machine learning prediction model for the data that was collected.

Six models were identified and trained. These include; Multiple Linear Regression, Linear Regression, Elastic Net, K-Nearest Neighbours, Decision Tree Regressor and Gradient Boosting Regressor. The best of them was the Gradient Boosting Regressor.

5.2.3 To train the suitable machine learning algorithm to predict future blood pressure readings of an individual using their past blood pressure readings and activity data.

After evaluating six different algorithms, the Gradient Boosting Regressor emerged as the most optimal choice for the subsequent steps of the process. The identified algorithm was then employed to train and construct a predictive model specifically tailored for the estimation of individuals' BP. This involved a comprehensive process of training the model using relevant data, refining its parameters, and establishing an effective way for accurate BP prediction.

5.2.4 To validate the performance of the machine learning algorithm for learning and predicting future levels of an individual's blood pressure

Validation of the predictive model was the last objective and it involved confirming whether the model selected was accurate in the prediction of the dependent variables. The researcher took some participants who helped in the validation of the model by using the phone application and the smartphone for 2 weeks. The data collected then helped to validate the model. Using Gradient Boosting Regressor, the model was used to predict the participants BP using their calendar events quite accurately.

5.3 Conclusions

The purpose of this study was to develop a machine learning solution that can predict an individual’s future BP level based on their past BP readings and activity data. This section provides the conclusions of this study according to the objectives mentioned in section 1.5.

Objective One: In order to get the data needed for the prediction model, an application was developed that was used together with a smartwatch. This process involved the creation of the Smarthealth application which was then given to the individuals who volunteered to participate in the data collection process. The application together with the smartwatch were well received by the participants. Most of the individuals were concerned whenever the application or the watch was not working, and they therefore communicated their concerns to the researcher. Almost all the participants wanted to either remain with the watches, or buy them. They liked the functionality of the watch and the Smarthealth application. Some referred friends who then wanted to take part in the study. The users also stated that the application let them be aware of their moods, and therefore, correct their mood and calm themselves down whenever they found themselves in a bad mood.

The limitation encountered in this section was the time it took for the development of the Smarthealth application. This is because the watches and the application had to be tested several times with different smartwatches and smartphones in order to create an application that can be used optimally. Another limitation was that the Bluetooth connectivity between the smartwatch and the application at times was not perfect, so the participants had to manually take their measurements.

In conclusion, the Smarthealth application together with the smartwatch were able to aid in the collection of the data needed for the creation of the predictive model.

Objective Two: The researcher was able to investigate and identify the best machine learning model by testing different predictive models and finding the one that made the best predictions. Gradient Boosting Regression performed the best with a mean square error of 0.19 and an R^2 score of 0.99. The analyses also revealed strong relationship between some variables, examples being the BMI and sleep pattern, the BMI and age and between the medication and age of an individual to name a few.

The limitation encountered in this section was the lack of complete data, mainly because of the Bluetooth connectivity between the smartwatch and the application. Other limitations include; smartphone battery drained faster than usual, and the participants complained about wearing the smartwatches because they were not used to wearing anything on their wrists.

Objective Three: The Gradient Boosting Regressor algorithm emerged as the most optimal choice for designing a predictive model to estimate individuals' blood pressure. Despite the limitations related to incomplete data, as highlighted in item ii, the model effectively addressed this challenge by incorporating variables exhibiting significant correlation with the target variables. As a result, the GBR model successfully achieved the objective of forecasting future BP levels for individuals.

Objective Four: The performance of the model was validated using two metrics, the mean square error and R Squared (R^2). Both scoring methods gave good results of 0.19 and 0.99 respectively. The scores showed that the model achieved its main aim of predicting an individual’s future BP. Validation was also done by sampling 5 individuals who recorded their data for two weeks, the model then predicted their future BP accurately.

5.4 Recommendations

This section offers a comprehensive array of recommendations tailored to address the needs of both policymakers and practitioners. Policymakers will find guidance for evidence-based policy formulation, while practitioners will discover practical strategies for enhancing the effectiveness of their activities, fostering synergy between theory and practice.

5.4.1 Recommendations for Policy and Practice

Embrace data-driven decision making: Policymakers should acknowledge the importance of data-driven decision making and advocate for its incorporation into policy formulation and implementation procedures. This entails investing in capabilities for data collection, analysis, and interpretation in order to gain actionable insights.

Encourage collaboration among policymakers, researchers, and data scientists to capitalize on interdisciplinary skills: Policymakers may guarantee that policy decisions are well-informed, evidence-based, and successfully address complex societal concerns by building partnerships.

Encourage responsible AI adoption: Policymakers should address the ethical and societal ramifications of AI technology on a proactive basis. Striking a balance between AI innovation and ethical considerations is critical to realizing AI's full promise while reducing dangers and maintaining responsibility.

In the first objective which was to develop an application for data collection, when future smartwatches are created with better Bluetooth connectivity, it is recommended they be used together with the Smarthealth application that was built. This is because of the loss of data due to lack of connectivity between the smartphone and the built application. When the smartwatch connects well with the application, there will be even less data that is lost, this will therefore enable creation of an even better predictive model.

For the second objective which was to investigate a machine learning model, further studies could include the use of other models that can predict panel data. This is because there are very many models to choose from, panel data is wide and contains more information. Models that use this type of data can be able to provide meaningful information that time series data and cross-sectional data cannot produce.

Considering the third objective aimed at training the most effective machine learning algorithm identified during the investigation phase, it is strongly advised to extensively employ the Gradient Boosting Regressor as a predictive model, particularly within the medical domain. This recommendation stems from its notable precision in forecasting, highlighting the potential benefits it can offer in medical applications.

In terms of validation of the predictive model, which was the third objective, further studies could use alternative evaluation methods to further check the correctness of the models.

5.4.3 Recommendations for Further Research

This research involved the use of smartwatches, a smartphone application and a predictive machine learning model to be able to predict individuals' blood pressure. The research was able to achieve its purpose. The methods of this research can be applied in making custom informative notifications through a phone application, website, email or Short Message Service (SMS), for each individual in order to prevent high BP or even lower BP in case of a hypertensive patient.

REFERENCES

A. Dhanasekar, D. R. (2016). Analysis of Association Rule for Heart Disease Prediction from Large Datasets. *International Journal of Innovative Research in Science, Engineering and Technology*, 18059-18063. ADDIN ZOTERO_BIBL CSL_BIBLIOGRAPHY Abdul Rehman, A., & Alharthi, K. (2016). *An introduction to research paradigms*. *3*.

Hypertension. (2023, March 16). https://www.who.int/news-room/fact-sheets/detail/hypertension

Kitt, J., Fox, R., Tucker, K., & McManus, R. (2019). *New Approaches in Hypertension Management: A Review of Current and Developing Technologies and Their Potential Impact on Hypertension Care | SpringerLink*. https://link.springer.com/article/10.1007/s11906-019-0949-4

Merai, R., Siegel, C., Rakotz, M., Basch, P., Wright, J., Wong, B., & Thorpe, P. (2016). CDC Grand Rounds: A Public Health Approach to Detect and Control Hypertension. *Morbidity and Mortality Weekly Report*, *65*(45), 1261–1264.

Ulz, J. (2023, March 20). *What is a Research Paradigm? Types and Examples | Researcher. Life*. Researcher Life. https://researcher.life/blog/article/what-is-a-research-paradigm-types-examples/

Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Dennison Himmelfarb, C., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., MacLaughlin, E. J., Muntner, P., Ovbiagele, B., Smith, S. C., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., … Wright, J. T. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ ASH/ASPC/NMA /PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension*, *71*(6), 1269–1324. https://doi.org/10.1161/HYP.0000000000000066

Yen, H.-Y. (2021). Smart wearable devices as a psychological intervention for healthy lifestyle and quality of life: A randomized controlled trial. *Quality of Life Research*, *30*(3), 791–802. https://doi.org/10.1007/s11136-020-02680-6

Yen, H.-Y., & Huang, W.-H. (2022). The efficacy of commercial smartwatches with a blood pressure-monitoring feature: A pilot randomized controlled trial. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing*, *54*(3), 324–331. https://doi.org/10.1111/jnu.12740

Zhou, Z.-H. (2021). *Machine Learning*. Springer Nature.

Zierle-Ghosh, A., & Jan, A. (2023). Physiology, Body Mass Index. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK535456/

Abhishek Gupta, I. M. (2016, October 14). *5G*. Retrieved from EDN Network: https://www.edn.com/5G/4442859/The-basics-of-Bluetooth-Low-Energy--BLE--

Al-Masri, A. (2019, January 10). *How does back-propagation in artificial neural networks work*. Retrieved from Towards Data Science: https://towardsd atascience.co m/how-does-back-propagation-in-artificial-neural-networks-work-c7cad873ea7

Amina J Mohammed, T. A. (2018). Healthy living, well-being and the sustainable development goals. *PubMed Central*, 590.

André Henriksen, M. H. (2018). Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *Journal of Medical Intgernet Research*.

Aprameya, A. (2016, April 6). *Quota Sampling: When To Use It and How To Do It Correctly*. Retrieved from Humans of Data: https://humansofdata.atlan .com/2016/04/quota-sampling-when-to-use-how-to-do-correctly/

Beers, B. (2019, May 1). *Terms*. Retrieved from Investopedia: https://www.inve stopedia.com/terms/r/regression.asp

Bing Zhang, H. R. (2019). Predicting blood pressure from physiological index data using the SVR algorithm. *BMC Bioinformatics*.

Borgelt, C. (2005). An Implementation of the FP-growth Algorithm. *ACM*, 1-5.

Bradley Camburn, V. V. (2017). Design prototyping methods: state of the art in strategies, techniques, and guidelines. *Design Science*, 1-33.

Brownlee, J. (2020, August 15). *A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/

Chauhan, N. S. (2019, October 3). *Introduction to Artificial Neural Networks (ANN)*. Retrieved from Towards Data Science: https://towardsdatascience.com /introduction-to-artificial-neural-networks-ann-1aea15775ef9

Chayakrit Krittanawong, A. S. (2018). Future Direction for Using Artificial Intelligence to Predict. *Springer*.

Chengyin Ye, T. F.-M. (2018). Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. *Jourmal of Medical Internet Research*.

Cherry, K. (2019, May 5). *What is the Experimental Method*. Retrieved from Very Well Mind: https://www.verywellmind.com/what-is-the-experimental-method-2795175

Cirruse Salehnasab, F. J. (2014). Use association rules to study the relation between variables that affect high blood pressure. *Acta Healthmedica*.

Copeland, B. J. (2019, April 11). *Technology*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/technology/artificial-intelligence

CrackMBA. (2012, January 26). *Evolutionary Prototype Model*. Retrieved from CrackMBA: http://crackmba.com/evolutionary-prototype-model/

Daniel LaFreniere, F. Z. (2016). Using Machine Learning to Predict Hypertension. *IEEE*, (pp. 1-7).

Daniel LaFreniere, F. Z. (2016). Using Machine Learning to Predict Hypertension from a Clinical Dataset. *Symposium Series on Computational Intelligence (SSCI).* IEEE.

Data Robot. (n.d.). *Wiki: Prediction*. Retrieved from Data Robot: https://www .datarobot.c om/wiki/prediction/

Dormehl, L. (2019, May 1). What is an artificial neural network? Here's everything you need to know. *Digital Trends*, pp. https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/.

Doyle, A. (2019, April 30). *Health: Heart Disease; Top iPhone and Android apps*. Retrieved from Healthline: https://www.healthline.com/health/heart-disease/top-iphone-android-apps#instant-heart-rate

Enid Wai-Yung Kwong, H. W.-H. (2016). A prediction model of blood pressure for telemedicine. *Health Informatics Journal*, 227-244.

Ezzat, A. (2018). Challenges and Solutions in Drug-Target Interaction Prediction. *Research Gate*.

Fan Liu, X. Z. (2018). Identification of Hypertension by Mining Class Association Rules from Multi-dimensional Features. *International Conference on Pattern Recognition (ICPR)* (pp. 3114-3119). Beijing: IEEE.

FrankenfIield, J. (2018, March 9). *Artificial Neural Networks (ANN) Defined*. Retrieved from Investopedia: https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp

Ghahrai, A. (2018, December 2). *Rapid Application Development*. Retrieved from Testing Excellence: https://www.testingexcellence.com/rapid-application-development-rad/

Giuseppe Mancia, F. R. (2017). Blood pressure control in hypertension. Pros and cons of available treatment strategies. *Journal of Hypertension*, 225–233.

Guo, Y. (2017, September 1). *The 7 Steps of Machine Learning*. Retrieved from Towards Data Science: https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e

Harsh Kukreja, B. N. (2016). An Introduction to Artificial Neural Network. *International Journal Of Advance Research And Innovative Ideas In Education*, 27-30.

Hudson Fernandes Golino, L. S. (2014). Predicting Increased Blood Pressure Using Machine Learning. *Journal of Obesity*, 1-12.

Ioannis Kavakiotis, G. T. (2013). Mining Frequent Patterns and Association Rules from Biological Data. In A. Y. Mourad Elloumi, *Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data*. Hoboken: John Wiley and Sons.

Ivanna Baturynska, K. M. (2021). Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms. *Journal of Intelligent Manufacturing* , 179–200 .

Jain, A. (2018, April 09). *Articles*. Retrieved from DZone: https://dzone.com/articles/machinex-laymans-guide-to-association-rule-learnin

Jamie Kitt, R. F. (2019). New Approaches in Hypertension Management: A Review of Current and Developing Technologies and Their Potential Impact on Hypertension Care. *Springer Open Choice*.

Jan A. Staessen, Y. L. (2017). Blood Pressure Measurement Anno 2016. *American Journal of Hypertension*.

Janusz Kaczorowski, M. D. (2012). Measurements of Blood Pressure: New Developments and Challenges. *British Columbia Medical Journal*, 399-403.

K, D. (2019). *Implementing Gradient Boosting in Python*. Retrieved from Paperspace Blog: https://blog.paperspace.com/implementing-gradient-boosting-regression-python/#:~:text=Gradient%20boosting%20Regression%20calculates%20the,maps%20features%20to%20that%20residual.

Ken Peffers, T. T. (2007). A Design Science Research Methodology for. *Journal of Management Information Systems*, 45-77.

Kenya Medical Research Institute. (n.d.). *About KEMRI: Background*. Retrieved from KEMRI Web site: https://www.kemri.org/index.php/features-intro/background

Kenya National Bureau of Statistics. (2019, November 4). *2019 Population and Housing Census*. Retrieved from Kenya National Bureau of Statistics Web Site: https://www.knbs.or.ke/?page_id=3142

Kracher, C. M. (2019, January). *Professional Education*. Retrieved from dentalcare.com: https://www.dentalcare.com/en-us/professional-education/ce-courses/ce490/types-of-blood-pressure-devices

Krakoff, L. R. (2015). Blood Pressure Out of the Office: Its Time Has Finally Come. *American Journal of Hypertension*, 289–295.

Krzysztof Pytel, T. N. (2015). Anthropometric Predictors and Artificial Neural Networks in the diagnosis of Hypertension. *Federated Conference on Computer Science and Information Systems*, (pp. 287–290). Lodz.

Lawrence A. Palinkas, S. M. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Adm Policy Ment Health,* 533-544.

Leahy, P. (2019, January 16). *What Is Java*. Retrieved from ThoughtCo: https://www.thoughtco.com/what-is-java-2034117

MacGill, M. (2018, November 21). Articles. *Everything you need to know about hypertension*. Retrieved from Medical News Today: https://www.medicalnewstoday.com/articles/150109.php

Mahajan, S. K. (2018). IoT-Fog based Healthcare Framework to Identify and Control Hypertension Attack. *IEEE Internet of Things Journal*, 1-8.

Mancia, G. (2012). Short- and Long-Term Blood Pressure Variability. *American Heart Association Journals*, 512–517.

Masui, T. (2022). All You Need to Know about Gradient Boosting Algorithm − Part 1. Regression. *Towards Data Science*.

Masui, T. (2022, January 20). *All You Need to Know about Gradient Boosting Algorithm − Part 1. Regression*. Retrieved from Towards Data Science: https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502

Merai R, S. C. (2016). CDC Grand Rounds: A Public Health Approach to Detect and Control Hypertension. *MMWR Morb Mortal Wkly*, 1261-1264.

Middleton, F. (2020, 13 January). *Types of Validity*. Retrieved from Scribbr Web site: https://www.scribbr.com/methodology/types-of-validity/

Mohssen Mohammed, M. B. (2017). *Machine learning : algorithms and applications.* Boca Raton: CRC Press.

Moi Teaching and Referral Hospital. (n.d.). *About Us*. Retrieved from MTRH: http://www.mtrh.go.ke/?page_id=598

Muoio, D. (2018, December 20). Omron's smartwatch blood pressure monitor cleared by FDA, launches in January. *Mobi Health News*.

Mutuvi, S. (2019, April 16). *Introduction to Machine Learning Model Evaluation*. Retrieved from Heartbeat Fritz AI: https://heartbeat.fritz.ai/introduction-to-machine-learning-model-evaluation-fa859e1b2d7f

Nichols, H. (2017, November 16). *Articles*. Retrieved from Medical News Today: https://www.medicalnewstoday.com/articles/320095.php

Nichols, H. (2017, February 23). The top 10 leading causes of death in the United States. *Medical News Today*.

Nohria, R. (2017). Comparative study of Adaptive neuro-fuzzy and fuzzy inference system for diagnosis of hypertension. *IEEE: International Conference on Computing Methodologies and Communication (ICCMC)*, (pp. 406-411).

Olive M Mugenda, A. G. (2003). *Quantitative and qualitative approaches.* Nairobi: African Centre for Technology Studies (ACTS).

Omron Healthcare. (2019, January 16). Retrieved from Omron Healthcare: https://omronhealthcare.com/2019/01/omron-heartguide-first-wearable-blood-pressure-monitor-top-award-winner-ces-2019/

Ooko, S. (2019, February 26). Lifestyle changes cut need for hypertension medicine. *Business Daily*.

Pexels. (n.d.). *Search*. Retrieved from Pexels Web site: https://www.pexels.com/

Phukkhapan Ghoshachandra, C. L. (2017). oHealth: A Self-Care Android Application for Senior Citizens with Hypertension. *ICT International Student Project Conference (ICT-ISPC).* Nakhon Pathom: IEEE.

Rakesh Agrawal, T. I. (1993). Mining association rules between sets of items in large databases. *ACM Digital Library*, 207-216.

Reilly, C. (2018, June 4). *Reviews: Asus Vivowatch bp preview*. Retrieved from Cnet: https://www.cnet.com/reviews/asus-vivowatch-bp-preview/

Renuka Patnaik, M. C.-C. (2018). Predicting the occurrence of essential hypertension using annual health records. *Second International Conference on Advances in Electronics, Computer and Communications (ICAECC-2018).* IEEE.

Ronilda C. Lacson, B. B. (2018). Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients. *Clinical Kidney Journal*, 206–212.

Roßbach, P. (2018, October 2). *Neural Networks vs. Random Forests – Does it always have to be Deep Learning?* Retrieved from Frankfurt School: https://blog.frankfurt-school.de/neural-networks-vs-random-forests-does-it-always-have-to-be-deep-learning/

Runge Chen, Y. Y. (2017). 3-Year Risk Prediction of Coronary Heart Disease in Hypertension Patients: A Preliminary Study*. *IEEE*, (pp. 1182-1185).

Ryan, V. (2013). *despro_flsh*. Retrieved from Technology Student: http://www.technologystudent.com/despro_flsh/evalintegr1.html

Satyanarayana Nimmala, Y. R. (2018). High blood pressure prediction based on AAA++ using machine-learning algorithms. *Cogent Engineering*, 1-12.

Shackelford, K. (2019, April 21). *Health Complications of Hypertension*. Retrieved from Very Well Health: https://www.verywellhealth.com/complications-of-hypertension-1763820

Sherrell, L. (2013). Evolutionary Prototyping. In *Encyclopedia of Sciences and Religions* (pp. 28-85). Dordrecht: Springer.

Shikha Singh, R. S. (2017). Prevalence and Associated Risk Factors of Hypertension: A Cross-Sectional Study in Urban Varanasi. *International Journal of Hypertension*, 1.

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics,*, 39-42.

Shukri F. Mohamed, M. K. (2018). Prevalence, awareness, treatment and control of hypertension and their determinants: results from a national survey in Kenya. *BMC Public Health*.

Silbert, S. (2019, April 24). *An Introduction to Smartwatches*. Retrieved from Lifewire: https://www.lifewire.com/an-introduction-to-smart-watches-3441381

Stefan Cronholm, G. G. (2003). Strategies for Information Systems Evaluation- Six Generic Types. *Electronic Journal of Information Systems Evaluation (EJISE)*, 1-13.

Tam Minh Tran, N. M. (2014). Changes in blood pressure classification, blood pressure goals and pharmacological treatment of essential hypertension in medical guidelines from 2003 to 2013. *ScienceDirect*.

Thomas, M. (2019, April 18). *Introduction to Neural Networks: Advantages and Applications*. Retrieved from Marktechpost Web site: https://www.marktechpost.com/2019/04/18/introduction-to-neural-networks-advantages-and-applications/?cn-reloaded=1

Trochim, W. M. (2006, October 20). *Knowledge Base*. Retrieved from Social Research Methods: http://www.socialresearchmethods.net/kb/reltypes.php

Ugurlu, T. (2018, November 8). *Iterative and Incremental Software Development — A Formula 1 Analogy*. Retrieved from Medium Web Site: https://medium.com/ingeniouslysimple/iterative-and-incremental-software-development-a-formula-1-analogy-bd58cefc853d

United Nations. (2018). *The Sustainable Development Goals Report 2018.*

University of Southern California. (2020, March 3). *Theoretical Framework*. Retrieved from USC Libraries: https://libguides.usc.edu/writingguide/theoreticalframework

Walker, A. (2018, November 19). *IDE*. Retrieved from Learning Hub: https://learn.g2crowd.com/ide

World Health Organization. (2018). *The top 10 causes of death.*

World Health Organization. (2019, September 13). *Hypertension*. Retrieved from WHO Web Site: https://www.who.int/news-room/fact-sheets/detail/hypertension

World Health Organization. (n.d). *Raised blood pressure.*

Xavier Ducrohet, T. N. (2013, May 15). Android Developers Blog. *Android Studio: An IDE built for Android*.

Yazdi, D. (2019, January 23). Opinion: Blood pressure measuring smartwatch is a paradigm shift for hypertension management. *Mobi Health News*. Retrieved from Mobi Health News.

APPENDICES

Appendix I: Questionnaire

a. Data that will be collected from the participants only once

1. Name ______________________
2. Gender

   o Male o Female
3. Date of Birth ______________________
4. Height ______________________
5. Weight ______________________
6. Family history of hypertension

   o Male o Female

b. Data collection from participants through the smartwatch and the smartphone application

1. Systolic blood pressure ______________________
2. Diastolic blood pressure ______________________
3. Smoking status
   a. o Yes o No
4. Exercise level ______________________
5. Alcohol level ______________________
6. Activity data ______________________
7. Future calendar events ______________________
8. Are you currently receiving drugs forany medical condition?
   a. o Yes o No

Appendix II: Sample Code

```
In [1]: import numpy as np
        import pandas as pd
        import scipy
        import sklearn as sk
        import matplotlib.pyplot as plt
        import seaborn as sb

        from sklearn.linear_model import LogisticRegression
        from sklearn.preprocessing import LabelEncoder

        from sklearn.feature_selection import RFE        #recursive feature elimination
        from sklearn.model_selection import train_test_split
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import KFold
        from sklearn.pipeline import Pipeline
        from sklearn.preprocessing import StandardScaler

        from sklearn.linear_model import LinearRegression
        from sklearn.linear_model import Lasso
        from sklearn.linear_model import ElasticNet
        from sklearn.tree import DecisionTreeRegressor
        from sklearn.neighbors import KNeighborsRegressor
        from sklearn.ensemble import GradientBoostingRegressor

        from sklearn.model_selection import GridSearchCV
        from sklearn.metrics import mean_squared_error

        %matplotlib inline

In [2]: df = pd.read_csv('Users_Data_U.csv')

In [3]: sb.pairplot(df)
        sb.set(rc = {'figure.figsize':(15,8)})
        plt.savefig('pairplot.png')
...

In [4]: # converting type of columns to 'category'
        df['User'] = df['User'].astype('category')

        # label encoding using 'category codes' approach
        df['User'] = df['User'].cat.codes

        df.info()
...

In [5]: print(df["Age"].mean())

        47.915336804972284

In [6]: # convert to datetime
        df['Hour'] = pd.to_datetime(df['Time'])

In [7]: # convert to hour
        df['Hour'] = df['Hour'].dt.hour

In [8]: df.head(20)
...

In [9]: df.info()
...

In [10]: # feature engineering
         df.Gender = df.Gender.astype(str)
         df.Hyper_hist = df.Hyper_hist.astype(str)
         df.Activity = df.Activity.astype(str)
         df.Mood = df.Mood.astype(str)

In [11]: # Using Labelencoder to convert categorical to numerical
         le = LabelEncoder()
         df['Gender'] = le.fit_transform(df['Gender'])
         df['Hyper_hist'] = le.fit_transform(df['Hyper_hist'])
         df['Activity'] = le.fit_transform(df['Activity'])
         df['Mood'] = le.fit_transform(df['Mood'])

In [12]: df = df.set_index('User')

In [13]: df.head(10)
...

In [14]: # picking out the relevant attributes for regression modelling
         correlation = df.corr(method='pearson')
         columns = correlation.nlargest(10, 'Sbp').index
         columns

Out[14]: Index(['Sbp', 'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'Hour', 'Mood',
                'Age', 'Gender'],
               dtype='object')
```

Appendix III: Ethical Clearance letter

KABARAK UNIVERSITY

BOARD OF POST GRADUATE STUDIES

Private Bag - 20157
KABARAK, KENYA
http://kabarak.ac.ke/institute-postgraduate-studies/

Tel: 0773265999
E-mail: directorpostgraduate@kabarak.ac.ke

15th July, 2020

The Director General
National Commission for Science, Technology & Innovation (NACOSTI)
P.O. Box 30623 – 00100
NAIROBI

Dear Sir/Madam,

RE: DAISY JEPKOECH KIPTOO- GMI/NE/0386/05/17

The above named is a candidate at Kabarak University pursuing Master's degree in Information Technology. He is carrying out a research entitled ***"An Artificial Neural Network Model for the Prediction of an Individual's Short-term Blood Pressure"***. She has defended her proposal and has been authorised to proceed with field research.

The information obtained in the course of this research will be used for academic purposes only and will be treated with utmost confidentiality.

Please provide the student with a research permit to enable her to undertake the research.

Thank you.

Yours faithfully,

KABARAK UNIVERSITY
DIRECTOR
POST GRADUATE STUDIES
15 JUL 2020
PRIVATE BAG -20157, KABARAK

Dr. Wilson O. Shitandi
DIRECTOR, INSTITUTE OF POST GRADUATE STUDIES

Kabarak University Moral Code
As members of Kabarak University family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord. (1 Peter 3:15)

Kabarak University is ISO 9001:2015 Certified

Appendix IV: NACOSTI Research Permit

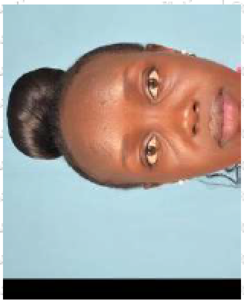
REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 153713

Date of Issue: 21/July/2020

RESEARCH LICENSE



This is to Certify that Miss.. Daisy Jepkoech Kiptoo of Kabarak University, has been licensed to conduct research in Uasin-Gishu on the topic: AN ARTIFICIAL NEURAL NETWORK MODEL FOR THE PREDICTION OF AN INDIVIDUAL'S SHORT TERM BLOOD PRESSURE for the period ending : 21/July/2021.

License No: NACOSTI/P/20/5893

153713

Applicant Identification Number

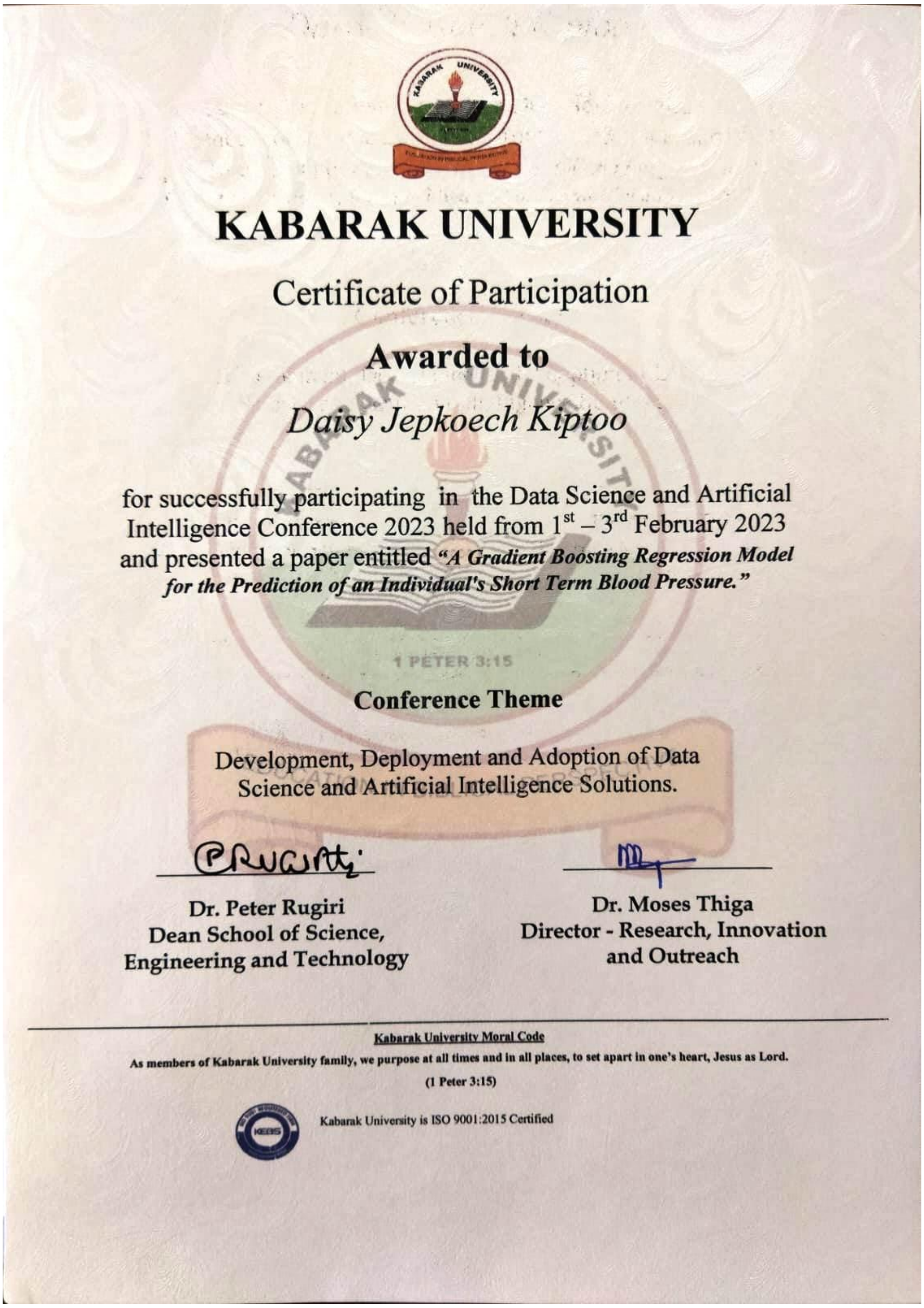
Director General

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix V: Evidence of Conference Participation



KABARAK UNIVERSITY

Certificate of Participation

Awarded to

Daisy Jepkoech Kiptoo

for successfully participating in the Data Science and Artificial Intelligence Conference 2023 held from 1st – 3rd February 2023 and presented a paper entitled ***"A Gradient Boosting Regression Model for the Prediction of an Individual's Short Term Blood Pressure."***

Conference Theme

Development, Deployment and Adoption of Data Science and Artificial Intelligence Solutions.

PRugiri

Dr. Peter Rugiri
Dean School of Science, Engineering and Technology

Dr. Moses Thiga
Director - Research, Innovation and Outreach

Kabarak University Moral Code

As members of Kabarak University family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord.

(1 Peter 3:15)

Kabarak University is ISO 9001:2015 Certified

Appendix VI: List of Publication

KABARAK JOURNAL OF RESEARCH & INNOVATION

Private Bag - 20157
KABARAK, KENYA
Email: editorial@kabarak.ac.ke

Tel: 254-51-343234/5
Fax: 254-051-343529
www.kabarak.ac.ke

OUR REF: KABU01/JISEA/07/07/01

06th December, 2023

Dear D. Kiptoo,

SUBJECT: PAPER ACCEPTANCE

We are pleased to let you know that your submission to the Journal of Information Systems in Eastern Africa (JISEA) has been accepted for publication. Details of the submission are as follows:

TITLE

AN EVALUATION AND SELECTION OF MACHINE LEARNING MODELS FOR BLOOD PRESSURE PREDICTION

AUTHORS

Daisy Kiptoo, Moses Thiga, Peter Rugiri, and Pamela Kimeto

ISSUE

No. 1(2023)

VOLUME

Vol. 1

Congratulations on this achievement and thank you so much for choosing JISEA.

Thank you.

Sincerely,

Dr. Michael N. Walekhwa
Editor in Chief

As members of Kabarak University family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord. (1 Peter 3:15)
Kabarak University is ISO 9001:2015 Certified