KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

HYPERTENSION PREDICTIVE MODEL WITH A NEURAL NETWORK APPROACH: A CASE STUDY OF KUMASI METROPOLIS

KNUST

By

Nina Ampah Bainn (BSc. Mathematics)

A thesis submitted to the Department of Mathematics,

Kwame Nkrumah University of Science and Technology in partial fulfilment of the requirements for the degree of

MASTER OF PHILOSOPHY

Faculty of physical Science

College of Science

December 2012

CERTIFICATION

I hereby declare that this submission is my work towards the Mphil and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgment has been made in the text.

KNUST

Nina Ampah Bainn (PG 5073910) …………………..... …………………

Student Name & ID Signature Date

Certified by:

E. Owusu- Ansah …………………... ……………………..

Supervisor(s) Name Signature Date

NWSAPƆ WƆSANE NO BADWENMA

Certified by:

Mr K. F Darkwah ……………………… ………………………

Head of Dept. Name Signature Date

DEDICATION

This work is dedicated to my family for their love and support they gave during the writing of this thesis.

KNUST

ABSTRACT

Most countries face high and increasing rates of Cardiovascular Disease. Hypertension is a major risk factor for many cardiovascular diseases the number one killer disease in Ghana today and a leading cause of deaths in adults. The objective of this study is to identify differences in the measurements recorded between the group that develop hypertension and the group that does not develop it. Secondly, develop a predictive mathematical model to estimate whether a person will develop hypertension using Artificial Neural Networks (ANN) from the Data and Tracies Data Exploration Studio (Software). ANN Feed Forwarding with Back-propagation methodology was adopted to extract significant patterns from a dataset containing 1027 observations. Using, Age, Body Mass Index, Systolic and Diastolic Blood Pressures as input descriptors. The data used for this study was collected by Healthy life Education, in the Kumasi metropolis, Ashanti Region of Ghana, West Africa. The findings of this study revealed that, the model built from ANN during a designed experiment, using different iterations gave the overall best model accuracy of 97.5% with a specificity of 98.2% and a sensitivity of 92.9%, showing that 16 people will be hypertensive. In conclusion, the study showed that the age group now for hypertensive state was below 50 years being overweight and pre-hypertensive. It was also predicted that 1.85% people will be hypertensive; hence increasing the state of hypertension. This means Artificial Neural networks techniques can be used efficiently to predict hypertension cases, therefore the outcome of this study can be used as an assistant tool by cardiologists to help them to make more consistent diagnosis for hypertension.

ACKNOWLEDGEMENT

I thank the Almighty God for bringing me this far in my pursuit of academic excellence. I thank my family for their love, support and prayers throughout my studies.

I also wish to thank E. Owusu-Ansah of the Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology (KNUST) for his corrections and guidance in ensuring successful completion this study.

In addition, I thank the management of Healthy Life Education KNUST, Kumasi, especially Dr. F.K.N Arthur of the Biochemistry department KNUST for allowing me carry out this research using Healthy Life Education research data.

My heartfelt gratitude again goes to E. Owusu-Ansah, Dr F.K.N Arthur and Dr. Ellis Owusu- Dabo of College of Health for perusing through my work.

My final appreciation goes to all those who have contributed in diverse ways in ensuring the successful completion of this work. God richly bless you.

KNUST

TABLE OF CONTENTS

CONTENTS..PAGE

KNUST

KNUST

LIST OF TABLES

LIST OF FIGURES

KNUST

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Hypertension is the single most important risk factor for cardiovascular disease and this remains a global public health challenge. Hypertension is the term used to describe high blood pressure (BP) which is the measurement of the force against the arteries as the heart pumps blood through the body. Hypertension or high blood pressure is a cardiac chronic medical condition in which the systemic arterial blood pressure is elevated.

According to PubMed Health (2011), Hypertension is classified as either primary (essential) hypertension or secondary hypertension; About 90% to 95% of cases are categorized as primary hypertension, which means high blood pressure with no obvious medical cause, the remaining 5% to 10% of cases (Secondary hypertension) are caused by other conditions that affect the kidneys, arteries, heart or endocrine system. Blood pressure is usually classified based on the systolic and diastolic blood pressures. Systolic blood pressure is the blood pressure in vessels during a heartbeat. Diastolic blood pressure is the pressure between heartbeats. A systolic or the diastolic blood pressure which measures a higher than the accepted normal values for the age of the individual is classified as pre-hypertension (more likely to develop high blood pressure) or hypertension. The BP readings are measured in millimeters of mercury (mmHg) with a sphygmomanometer and usually given as two numbers such as 120 over 80 (written as

120/80mmHg). Here the top number is the systolic pressure and the bottom number is the diastolic pressure. Systolic pressure is considered high if is140mmHg and over and normal if is below 120mmHg. Again diastolic pressure is rated high if is 90mmHg and normal if is below 80mmHg. In pre-hypertension cases systolic pressure may be between 120 and 139(mmHg) and diastolic pressure at 80 to 89(mmHg).

1.1.1 Classification and Prevention

Hypertension has several sub-classifications including, hypertension stage I, hypertension stage II .Other forms of hypertension are;

- **Neurogenic hypertension**- an alteration in neural cardiovascular.
- **Essential hypertension** is the most prevalent hypertension type, affecting 90%–95% of hypertensive Individuals and has no direct cause.
- **Accelerated hypertension** is associated with headache, vision disorders, nausea, and vomiting. These symptoms are collectively called hypertensive encephalopathy and are caused by severe small blood vessel congestion and brain swelling which is reversible if blood pressure is lowered.
- **Secondary hypertension**; high blood pressure that is caused by another medical condition or medication such as alcohol abuse, diabetes, birth pills, appetite suppressant, migraine medications etc.
- **Exercise hypertension** is when there is excessive high elevation in blood pressure during exercise. The range considered normal for systolic values during exercise is between 200 and 230 mm Hg. Exercise hypertension may indicate that an individual is at risk for developing hypertension at rest.

- **White coat hypertension** is when a person's blood pressure is normal when taken at home but is high when taken by the doctor (who may be wearing a white coat). This phenomenon is thought to be due to stress of seeing the doctor.

Hypertension can be prevented by current blood pressure level, sodium/potassium balance, detection and omission of environmental toxins, changes in organs (retina, kidney and heart among others), risk factors for cardiovascular diseases and the age at diagnosis. Following this, lifestyle changes are recommended to lower blood pressure, before the initiation of prescription drug therapy. The treatment for hypertension is the same as the recommended preventative lifestyle changes which includes dietary changes, physical exercise and weight loss. These have all been shown to significantly reduce blood pressure in people with hypertension. Some results of hypertension are as follows; Aortic dissection, brain damage, Silent Stoke, retina and kidney damages.

1.1.2 Mathematical Epidemic Modeling

Artificial Neural Networks (ANN) is currently a hot research area in medicine and it is believed that they will receive extensive application to biomedical systems in the next few years. At the moment, the research is mostly on modeling parts of the human body and recognizing diseases from various scans (e.g. cardiograms, CAT scans, ultrasonic scans, etc.). Neural networks are ideal in recognizing diseases using scans since there is no need to provide a specific algorithm on how to identify the disease. Neural networks learn by example so the details of how to recognize the disease are not needed. What is needed is a set of examples that are representative of all the variations of the disease.

The quantity of examples is not as important as the quantity. The examples need to be selected very carefully if the system is to perform reliably and efficiently.

Neural Networks are used experimentally to model the human cardiovascular system. Diagnosis can be achieved by building a model of the cardiovascular system of an individual and comparing it with the real time physiological measurements taken from the patient. If this routine is carried out regularly, potential harmful medical conditions can be detected at an early stage and thus make the process of combating the disease much easier. A model of an individual's cardiovascular system must mimic the relationship among physiological variables (i.e., heart rate, systolic and diastolic blood pressures, and breathing rate) at different physical activity levels. If a model is adapted to an individual, then it becomes a model of the physical condition of that individual. The simulator will have to be able to adapt to the features of any individual without the supervision of an expert. This calls for a artificial neural network which will be used in this thesis in determining the rate of hypertension in non- hypertensive states using four inputs; Body Mass Index, Age, Systolic and Diastolic Blood pressures.

1.1.3 Study Area

The study was carried out in Kumasi, Ashanti Region of Ghana. Ghana is located on the Atlantic Coast of West Africa, 4˚ north of the Equator. It occupies a land area of 238,537 km^2 and had a population of 23.5 million of which 56% is rural. Life expectancy at birth was 55 years for males and 60 for females in 2000. The country is divided into ten regions and 170 districts. Per capita total health expenditure as a percentage of GDP was 4.5% in 2003. The two largest cities Accra being the national and Kumasi, the regional capital of the Ashanti Region have the highest concentration of doctors in the country. About 60% of persons who report ill or injured consult a health practitioner while 32% purchase medicine directly from the pharmacy shops.

1.2 Statement of the Problem

KNUST

Hypertension is a major risk factor for many cardiovascular diseases. A comprehensive review of the prevalence of hypertension provides crucial information for the evaluation and implementation of appropriate programs.

Hypertension is the number one killer disease in Ghana today. Doctors at the Korle-Bu Teaching Hospital say almost 70% of all deaths at the hospital are caused by hypertensive conditions; hypertension is a silent killer because many have it for years without realizing it. It silently damages the brain, the heart, the kidneys and the eyes. The disease affects nearly one out of every five Ghanaian adults. A recent report on hypertension by the Ghana Health Service says more people are becoming hypertensive due to unhealthy lifestyles. Doctors explain that hypertension commonly referred to as BP is the major cause of strokes, heart attacks, heart failure and chronic renal failure. These and other blood pressure related diseases constitute more than half of all admission cases at Korle Bu teaching hospital. Hypertension is a major health care problem, in the urban centers its prevalence is about 30% and is the leading cause of deaths in adults in Kumasi and Accra.

1.3 Objectives of the Study

The objective of this study is to understand any general relationships between different patient characteristics and the propensity to develop hypertension, these include:

1. Identify differences in the measurements recorded between the group that develop hypertension and the group that does not develop it.

2. Develop a predictive mathematical model to estimate whether a patient will develop hypertension from normal through pre-hypertension using Artificial Neural Networks.

1.4 Methodology

This thesis employs a model based on biological information and from previous works. We present mathematical models that capture leading aspects of epidemiology of Hypertension in Ashanti region of Ghana. The model we decided to use in studying Hypertension is Artificial Neural Networks- Feed forwarding with Back -Propagation.

An application of a multi-layer feed forward network with a back propagation training algorithm is to learn an unknown function between input and output signals from the presentation of examples. It is hoped that the network is able to generalize correctly, so that input values which are not presented as learning patterns will result in correct output values. The assumption for this model is to make predictions or estimate using historical data to indicate people who will be hypertensive over time. An example is the work of Josin (1988), who used a two layer feed forward network with back propagation learning to perform the inverse kinematic transform which is needed by a robot arm controller.

1.5 Justification

Hypertension is clearly an important cardiovascular disease public health problem in Ghana, even in the poorest rural communities. Emphasizing on health promotion will aid in controlling this health issue. This thesis is intended to help find out the rate at which this disease increases using Artificial Neural Network to aid in detecting and diagnosing it early, since the occurrences of death from this disease generally affect the country's productivity and gross domestic product.

1.6 Thesis Organization

This thesis is organized into five main chapters. Chapter one presents the introduction of the thesis. This consists of the background of the study, the research statement of problem, objectives of the research, methodology, thesis justification and organization of the thesis. Chapter two is the literature review, which looks at work done by other researchers on the topic. Chapter three is the formulation of the mathematical model using ANN feed forwarding with back-propagation algorithm. Chapter four contains the data collection, analysis and results of ANN design experiment from the built model. Chapter five looks at summary, conclusions and recommendations.

KNUST

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

This chapter focuses on the review of related literature on the anthropology measurement of hypertension and Artificial Neural Network in disease model building. This comprises other related research findings, which will assist the researcher and serve as a guide for future research.

2.2 Related literature on cardiovascular disease

Many researchers have had the same or similar studies; therefore, the comparative studies of this chronic disease concentrating on it anthropometry and epidemiology. According to Whitworth (2003), hypertension is estimated to cause 4.5% of current global disease burden and is as prevalent in many developing countries, as in the developed world. Blood pressure-induced cardiovascular risk rises continuously across the whole blood pressure range. Countries vary widely in capacity for management of hypertension, but worldwide the majority of diagnosed hypertension is inadequately controlled. This statement addresses the ascertainment of overall cardiovascular risk to establish thresholds for initiation and goals of treatment, appropriate treatment strategies for non-drug and drug therapies, and cost-effectiveness of treatment. In conclusion it was stated that since publication of the WHO/ISH Guidelines for the Management of Hypertension in 1999, more evidence has become available to support a systolic blood pressure threshold of 140 mmHg for even low-risk patients. In high-risk patients there is evidence for lower thresholds. Lifestyle modification is recommended for all individuals.

The number of research studies conducted on cardiovascular disease in the past decades has increased dramatically, Chen- Huan (2000) research findings on the different association of hypertension and insulin-related metabolic syndrome between men and women in 8437 non-diabetic Chinese indicated that, a distinct insulin-resistance related metabolic syndrome characterized by hyperinsulinemia, dyslipidemia, and obesity was observed for men and women in this Chinese population. However, hypertension was linked to metabolic syndrome in women only. They first examined fasting insulin, glucose, triglyceride and high-density lipoprotein (HDL) – cholesterol levels, systolic blood pressure, body mass index, and waist-to-hip ratio in a dataset from 8437 non-diabetic residents (age range, 30 to 89 years) in Kinmen. Factors analysis in men shows 29.7% for anthropometrics and 29.4% in women, 18.1% for systolic blood pressure and glucose in men then 14.0% in women in all variables considered.

Twagirumukiza (2011) looked at the current and projected prevalence of the arterial hypertension in sub-Saharan African (SSA) by sex, age and habitat: an estimate from population studies. He stated that in sub-Saharan, data on hypertension prevalence in terms of urban or rural and sex difference are lacking, heterogeneous or contradictory in many countries underlying the need for national surveys and so there are no accurate estimate of hypertension burden. 17 studies pertaining to 11 countries were analyzed. The overall prevalence rate of hypertension in SSA for 2008 was estimated at 16.2% ranging from 10.6% in Ethiopia to 29.9% in Ghana. The estimated prevalence was 13.7% in rural areas, 20.7% in urban areas, 16.8% in males, and 15.7% in females. The total number of hypertensive in SSA was estimated at 75 million in 2008 and at

125.5million by 2025. This shows that the estimated number of hypertensive in 2008 is nearly four times higher than the last 2005 estimate of the World Health Organization Regional Office for Africa. Prevalence were significantly higher in urban than in rural populations.

Kunutsor (2009), research on descriptive epidemiology of blood pressure in a rural adult population in the Northern Ghana finding also shows, there are a limited data on blood pressure (BP) levels in rural populations in Sub-Saharan Africa. This is a cause for concern considering high BP contributes to a substantial public health burden in this population. Blood pressure levels were measured in a representative rural sample from Kassena-Nankana District of the Northern Ghana, West Africa and associations with anthropometric indices, age, sex and time of BP measurement where assessed. This cross-sectional survey was made on a random sample drawn from a population register and included 207 males and 367 females aged between 18 and 65 years. Overall prevalence of casual high BP was 19.3%. A regression analysis shows that systolic BP was significantly associated with age, BMI waist circumference and time of BP measurement. In multiple regression analyses these anthropometrics were independently associated with systolic and diastolic BP in women, while time of BP measurement was independently associated with systolic BP in men only. There was no increase in BMI with age and repeated measurements in 89 subjects yielded a regression dilution factor of 0.57 for systolic BP. Salt consumption per individual was estimated as equivalent to 12.5g of pure salt/day. He then concluded that the average BPs are not notably high in this predominantly lean rural West African population, but they do however increase as expected with age and BMI. Population-wide approaches need to be developed, appropriate to the level of medical provision, in order to address vascular disease risks resulting from higher than optimal BPs such as programs to reduce salt consumption.

Another researcher wanted to know whether it was essential to identify the best anthropometric index in any population to predict chronic disease risk. Esmaillzedeh (2004) findings on waist-to-hip ratio (WhpR) is a better screening measure for cardiovascular risk factors then other anthropometric indicators in Tehranian adult men. Taking a sample of 4,449 men aged 18-74, participants of the Tehran Lipid and Glucose study. Demographic data were collected; anthropometric indices (BMI, WhpR and waist circumference (WC)) and blood pressure were measured according to standard protocol. The presence of at least one risk factor from the three major cardiovascular risk factors (hypertension, dyslipidermia and diabetes) was also evaluated. Although all anthropometric indicators had a significant association to cardiovascular risk factors in all ages categories, the highest odds ratios were pertained to WhpR. Of the four individual indicators, WhpR had the highest sensitivity, specificity and accuracy to prediction cardiovascular risk factors. Cutoff points for WhpR were seen to have a higher percentage of correction prediction than BMI, WC and WhpR in all age categories. It was then concluded that WhpR is a better predictor for cardiovascular risk factors then the other anthropometric indicators.

According to the Ghana News Agency, GNA (2011), Statistics from Internal Medicine Specialists at the Komfo Anokye Teaching Hospital (KATH) and the Kwame Nkrumah University of Science and Technology (KNUST), shows disturbing upsurge of cases of hypertension and diabetes of cases of hypertension and diabetes in Kumasi and its environs. KATH has been seeing 120 new cases of diabetes every week. Again, 25 per cent of deaths at the referral facility are attributed to the abnormally high blood pressure disease and its related infections. KNUST also presented a paper on the hypertension situation, saying 28.7 per cent of people in the Kumasi Metropolis and its outlying communities have the disease and it is one of the leading causes of pre-mature deaths and disabilities and also account for 62 per cent of all strokes cases. Hence, they commended KATH for the initiative and urged them to use the knowledge acquired to improve the management of these disease and appealed to doctors and health officials to closely monitor the progress of treatment of their patients.

Addo (2006) Objective was to determine the prevalence, distribution and risk factors of hypertension among rural residence in Ghana. A cross sectional study in four rural communities in the Ga district of Ghana with all adults aged 18years and above. In her studies there were 362 subjects with mean age of 42.4. The prevalence of hypertension was 25.4%. She also stated that the adjusted odds ratio for hypertension with respect to smoking, alcohol consumption, job-related physical activity, family history, education, occupation and diabetes status did not attain statistical significance. She then concluded that, hypertension is now of public health significance in rural Ga district of Ghana. The high rate of hypertension was associated with low levels of awareness, drug treatment, and blood pressure control. Over weight and obesity are modifiable risk factors for hypertension that can be addressed through lifestyle interventions and most importantly integrating hypertension care into primary care in rural health facilities may prove beneficial.

Another important finding by Francesco (2008) and other research group in the UK on body size and blood pressure: an analysis of Africans and the African Diaspora indicates that blood pressure and BMI levels vary among populations of the African Diaspora. The effect of BMI on blood pressure levels diminishes as BMI increases. These results suggest a complex relationship between excess body weight, adiposity and energy expenditure. Blood pressure is directly and causally associated with BMI in populations worldwide. However, the relationship may vary across BMI of African origin. The relationship comparison between blood pressure and BMI was made in the population of African origins from Africa, the Caribbean, UK and USA. The result shows a positive relationship between both systolic and diastolic blood pressure and BMI. In men the slope for systolic BP varied from 0.27mm Hg per kg/m in the US to 1.72 in Ghana (Kumasi), in women the slope varied from 0.08 in South Africa to 1.32 in Republic of Congo. Similar variation in trends was seen for diastolic BP. The higher the BMI, shallower the slope hence no difference was seen after excluding persons who were being treated for hypertension.

A similar studies was conducted by Bosu (2010), he studied the epidemic of hypertension in Ghana: a systematic review. In this review he looked at hypertension being a major risk factor for many cardiovascular diseases in developing countries that is a comprehensive review of the prevalence of hypertension provides crucial information for evaluation and implementation of appropriate programmes based on hypertensive Adult in Ghana between 1970 and August 2009. Fifteen unique population-based article in non-pregnant humans were obtained also two relevant unpublished graduate studies theses from one university department were identified after a search of its 1996-2008 these. Their results shows that from all angles of anthropometric risk factors, sex difference were generally minimal whereas urban populations tends to have higher prevalence than rural population in studies with mixed population types. Factors independently associated with hypertension included older age group, over-nutrition and alcohol consumption. Whereas there was a trend towards improved awareness, treatment and control between 1972 and 2005, less the one third of hypertensive subjects were aware they had it and less than one-tenth had their BP controlled in most studies. He then concluded that hypertension is clearly an important public health problem in Ghana, even in the poorest rural communities. Emerging opportunities such as the National Health Insurance Scheme, a new health policy emphasizing health promotion and healthier lifestyles and effective treatment should help prevent and control hypertension.

2.3 Modeling using Neural Networks

Designing a predictive model for heart disease detection using data mining techniques was done by Damtew (2011). The general objective was to design a predictive model for heart disease detection using data mining techniques from Transthoracic Echocardiography Report dataset that is capable of enhancing the reliability of heart disease diagnosis using echocardiography. His methods were Knowledge Discovery in Database (KDD) methodology consisting of nine iterative and interactive steps was adopted to extract significant patterns from a dataset containing 7,339 echocardiography examination reports of patients. The data used for this study was collected by International Cardiovascular Hospital from October, 2008 to March, 2011. The findings of this study revealed all the models built from J48 Decision Tree classifier, Naïve Bayes classifier and Neural Network have high classification accuracy and are generally comparable in predicting heart disease cases. However, comparison that is based on True Positive Rate suggests that the J48 model performs slightly better in predicting heart disease with classification accuracy of 95.56%. In conclusion, his studies showed that data mining techniques can be used efficiently to model and predict heart disease cases.

In Ohno-Machado (1996) Medical applications of artificial neural networks: Connectionist models of survival stated that, although neural networks have been applied to medical problems in recent years, their applicability has been limited for a variety of reasons. One of those barriers has been the problem of recognizing rare categories. In her dissertation, she demonstrate, and prove the utility of, a new method for tackling this problem that is she developed a method that allows the recognition of rare categories with high sensitivity and specificity, and will show that it is practical and robust. This method involves the construction of sequential neural networks. Rare categories occur and must be learned if practical application of neural-network technology is to be achieved. Survival analysis is one area in which this problem appears. She tested the hypotheses that (1) sequential systems of neural networks produce results that are more accurate (in terms of calibration and resolution) than nonhierarchical neural networks; and (2) in certain circumstances, sequential neural networks produce more accurate estimates of survival time than Cox proportional hazards and logistic regression models. She used two sets of data to test the hypotheses: (1) a data set of HIV+ patients (AIDS Time-Oriented Health Outcome Study ATHOS data set); and (2) a data set of patients followed prospectively for the development of cardiac conditions (Framingham data set). Using the ATHOS data set saying that, a neural network model can predict death due to AIDS more accurately than a Cox proportional hazards model. Furthermore, she showed that a sequential neural network model is more accurate than a standard neural network model. Using the Framingham data set, shows that the predictions of logistic regression and neural networks are not significantly different, but that any of these models used sequentially is more accurate than its standard counterpart. The sequential use of predictive models for survival analysis is advantageous because it makes better use of the available information. It often increases resolution with no sacrifice of calibration. It also helps to delineate patterns of this ease progression for individuals, rather than for groups of patients.

KNUST

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter looks into the methodology using Artificial Neural Network (ANN) to model the cardiovascular disease. The method used in this thesis is feed forward network with back propagation learning. ANN an is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.

3.1.1 Why Neural Networks?

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques such as;

- Classification: pattern recognition, feature extraction and image matching.
- Noise reduction: Recognized patterns in inputs and produces noiseless output.
- Prediction: Extrapolation based on historical data.

3.2 The Mathematical Model

A neural network is a mathematical model that makes predictions based on a series of input descriptor variables. Like all prediction models, it uses a training set of examples to generate the model. This training set is used to generalize the relationships between the input descriptor variables and the output response variables. Once a neural network has been created, it can then be used to make predictions and are presented in optimization form.

3.2.1 Feed forward Network and Back-propagation Learning

Feed-forward ANNs (figure 3.1) allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs.

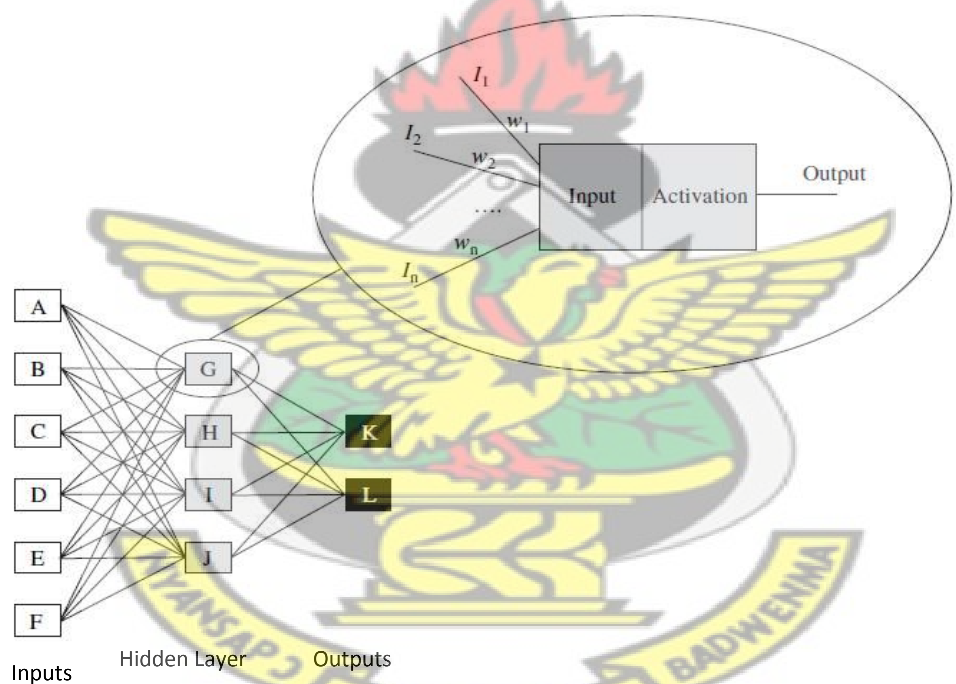


Figure 3.1: Feed forward Network

Source: Making Sense of Data, pg 181.

From the inputs to outputs, A to F are the Input node, G to J are the hidden layer nodes and K,L are the output nodes. A Node is a decision point within a decision tree and a point at which connections join within a neural network. Each node in the neural network calculates a single output value based on a set of input values. Activation function is a function that generates the output for the node. The commonest type of artificial neural network consists of three groups, or layers, of units: a layer of input units is connected to a layer of hidden units, which is connected to a layer of output units.

1. The activity of the input units represents the raw information that is fed into the network.
2. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units.
3. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

Learning process is memorisation of patterns and the subsequent response of the network.

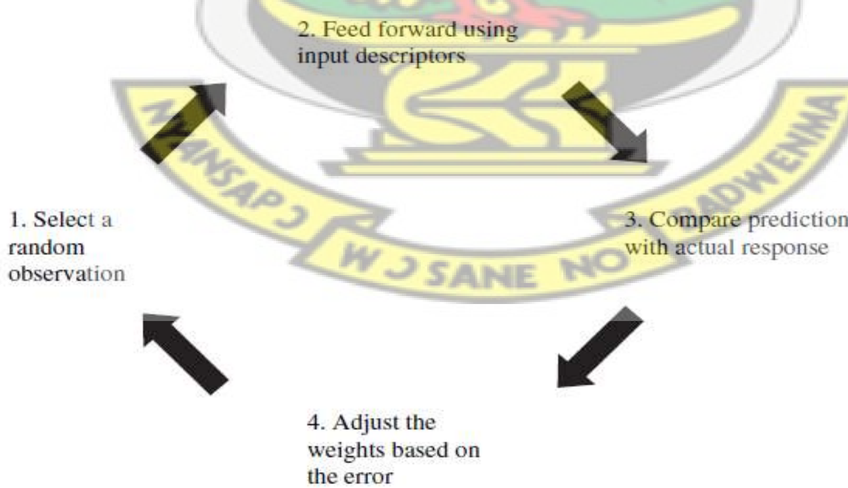


Figure 3.2: Learning process in Neural Network

Source: Making Sense of Data, pg 192.

Back propagation is the most common method of obtaining the many weights in the network. The basic back propagation algorithm is based on minimizing the error of the network using the derivatives of the error function.

3.3 The Back-Propagation Algorithm

In order to train a neural network to perform some task, we must adjust the weights of each unit in such a way that the error between the desired output and the actual output is reduced. This process requires that the neural network compute the error derivative of the weights (**EW**).

The algorithm computes each **EW** by first computing the **EA**, the rate at which the error changes as the activity level of a node is changed. For output units, the **EA** is simply the difference between the actual and the desired output. To compute the **EA** for a hidden unit in the layer just before the output layer, we first identify all the weights between that hidden unit and the output units to which it is connected. The **EW** is the product of the EA and the activity through the incoming connection. Before back-propagating, the **EA** must be converted into the **EI**, the rate at which the error changes as the total input received by a unit is changed.

3.3.1 The back-propagation Algorithm - a Mathematical Approach

Nodes are connected to one another. There is a real number associated with each connection, which is called the weight of the connection. We denote by *Wij* the weight of the connection. Two types of connection are usually distinguished: excitatory and inhibitory. A positive weight represents an excitatory connection whereas a negative weight represents an inhibitory connection. The pattern of connectivity characterises the architecture of the network.

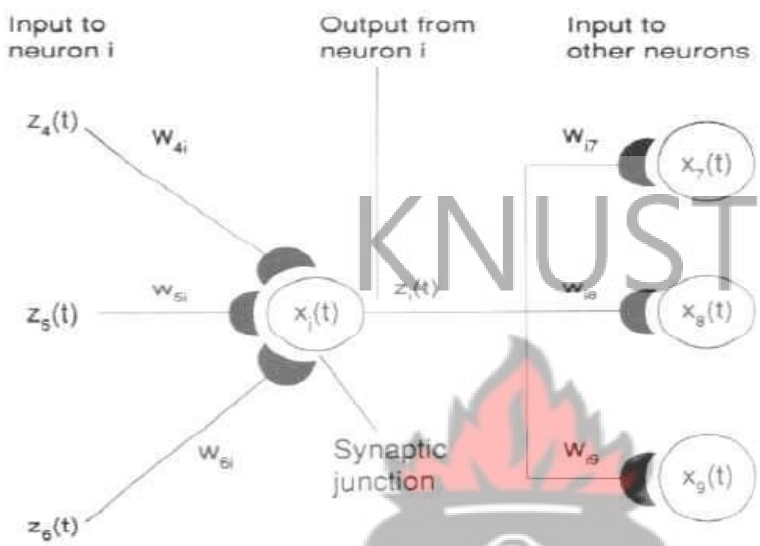


Figure: 3.3. Abstract Neuron

A node in the output layer determines its activity by following a two step procedure.

First, it computes the total weighted input x_j, using the equation:

$$x_j = \sum_i y_i W_{ij} \tag{3.10}$$

Where, *yi* is the activity level of the *jth* node in the previous layer and *Wij* is the weight of the connection between the *ith* and the *jth* node.

Next, the node calculates the activity y_j using some function of the total weighted input. Typically we use the sigmoid function:

$$y_j = \frac{1}{1+e^{-x_j}} \tag{3.11}$$

Once the activities of all output nodes have been determined, the network computes the error E, which is defined by the expression:

$$E = \frac{1}{2} \sum_{i} (y_i - d_i)^2 \tag{3.12}$$

Where, *yj* is the activity level of the jth node in the top layer and *dj* is the desired output of the jth node.

KNUST

The back-propagation algorithm consists of four steps:

1. Compute how fast the error changes as the activity of an output node is changed. This error derivative (EA) is the difference between the actual and the desired activity.

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j \tag{3.13}$$

2. Compute how fast the error changes as the total input received by an output node is changed. This quantity (EI) is the answer from equation (3.13) multiplied by the rate at which the output of a node changes as its total input is changed.

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{dy_j}{dx_j} = EA_j y_j (1 - y_j) \tag{3.14}$$

3. Compute how fast the error changes as a weight on the connection into an output node is changed. This quantity (EW) is the answer from equation (3.14) multiplied by the activity level of the node which the connection is derived from.

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial W_{ij}} = EI_j y_j \tag{3.15}$$

4. Compute how fast the error changes as the activity of a node in the previous layer is changed. This crucial step allows back propagation to be applied to multilayer networks. When the activity of a unit in the previous layer changes, it affects the activates of all the output nodes to which it is connected. So to compute the overall effect on the error, we add together all these separate effects on output nodes. But each effect is simple to calculate. It is the answer in equation (3.14) multiplied by the weight on the connection to that output unit.

$$\boldsymbol{EA_i} = \frac{\partial E}{\partial y_i} = \sum \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial \boldsymbol{y_i}} = \sum \boldsymbol{EI_j}\, \boldsymbol{W_{ij}} \qquad (3.16)$$

By using equation (3.14) and (3.16), we can convert the EAs of one layer of nodes into EAs for the previous layer. This procedure can be repeated to get the EAs for as many previous layers as desired. Once we know the EA of a node, we can use equation (3.14) and (3.15) to compute the EWs on its incoming connections.

3.4 Model Parameters

Most predictive models especially neural networks can be optimized by fine tuning different model parameters. This includes;

1. Number of cycles or epochs; One iteration through the process of providing the network with an input and updating the network's weights.
2. Hidden layers; both the number of hidden layers and the number of nodes in each hidden layer can influenced the quality of the results whether too many or few.
3. Learning rate; this influences how fast the neural networks learns.

4. Inputs/outputs; they are the descriptors or/ response variables and can be more than one.

3.5 Receiver Operating Characteristic (ROC) Curve in Prediction

The most popular measure of model fit in the cardiovascular literature has been the *c* statistic, a measure of discrimination also known as the area under the ROC curve, or the c index, its generalization for survival data. The ROC curve and its associated *c* statistic are functions of the sensitivity and specificity for each value of the measure or model. The sensitivity of a test is the probability of a positive test result, or of a value above a threshold, among those with disease (cases). The specificity is the probability of a negative test result, or a value below a threshold, among those without disease (non cases). It is commonly believed that sensitivity and specificity are properties of a test and are not subject to alteration by prevalence of disease, as are the positive and negative predicted values. This has been shown to be false, however, both theoretically and clinically. Both sensitivity and specificity can be influenced by case mix, disease severity, or risk factors for disease. For example, a test is likely to be more sensitive among more severe than among milder cases of disease. Similarly, specificity can depend on characteristics of non cases, such as gender, age, or prevalence of concomitant risk factors.

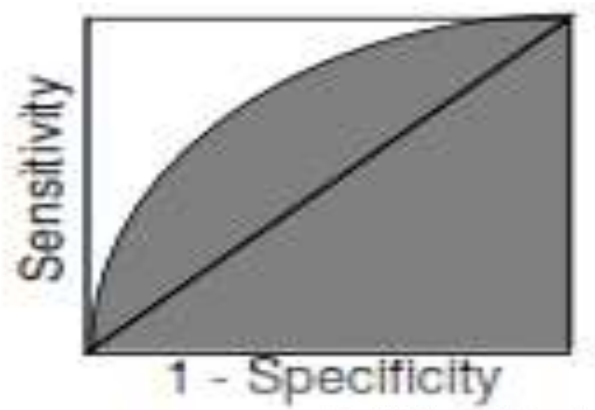


Figure: 3.4 The ROC curve

KNUST

The ROC curve is a plot of sensitivity versus 1−specificity (often called the false-positive rate) that offers a summary of sensitivity and specificity across a range of cut points for a continuous predictor. The area under the curve or *c* statistic, ranges from 0.5 (no discrimination) to a theoretical maximum of 1. Perfect discrimination corresponds to a *c* statistic of 1 and is achieved if the scores for all the cases are higher than those for all the non-cases, with no overlap. The *c* statistic is equivalent to the probability that the measure or predicted risk is higher for a case than for a non case. Note that the *c* statistic is not the probability that individuals are classified correctly or that a person with a high test score will eventually become a case. The latter is closer in meaning to the predictive value, or the probability of disease given the test result.

3.5.1 Artificial Neural Network Predictions

A neural network makes a prediction based on the input descriptor variables presented to the network and the weights associated with connections in the network. Figure 3.1 shows an observation presented to the network.

The quality of a classification model can be assessed by counting the number of correctly and incorrectly assigned observations. Jaccard distance is a method that handles only binary variables. The contingency table shown in Table 3.1 is used to calculate the Jaccard distance between two observations that have been measured over a series of binary variables where the actual response is compared against the predicted response for a binary variable.

- **Count:11** The number of observations that were true and predicted to be true (true positives).
- **Count:10** The number of observations that were false yet predicted to be true (false negatives)
- **Count:01** The number of observations that were true and predicted to be false (false positives).
- **Count:00** The number of observations that were false and predicted to be false (true negatives)

Table 3.1 shows relationships between two observations measured using a series of binary variables.

Table 3.1: Contingency table format

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Predicted Response
(Observation 1)</td></tr>
<tr><td>True(1)</td><td>False (0)</td></tr>
<tr><td rowspan="2">Actual Response
(Observation 2)</td><td>True (1)</td><td>$Count_{11}$</td><td>$Count_{01}$</td></tr>
<tr><td>False (0)</td><td>$Count_{10}$</td><td>$Count_{00}$</td></tr>
</table>

There are four calculations that are commonly used to assess the quality of a classification model:

1. Concordance: This is an overall measure of the accuracy of the model and is calculated with the formula:

$$\boldsymbol{Concordance} = \frac{(Count\ 11 + Count\ 00)}{(Count\ 11 + Count\ 10 + Count\ 01 + Count\ 10 + Count\ 00)} \qquad (3.17)$$

2. Error rate: This is an overall measure of the number of prediction errors and the formula is:

$$\boldsymbol{Erro\ Rate} = \frac{(Count\ 10 + Count\ 01)}{(Count 11 + Count 10 + Count\ 01 + Count\ 00\ )} \qquad (3.18)$$

3. Sensitivity: This is an assessment of how well the model is able to predict 'true' values and the formula is:

$$\boldsymbol{Sensitivity} = \frac{(Count\ 11)}{(Count\ 11 + Count\ 01)} \qquad (3.19)$$

4. Specificity: This is an assessment of how well the model is able to predict 'false' values and the formula is:

$$\boldsymbol{Specificity} = \frac{(Count\ 00)}{(Count\ 10 + Count\ 00)} \qquad (3.20)$$

3.6 Research Design

The study was carried out in Kumasi in the Ashanti Region of Ghana. The sample for the study was chosen purposively and limited to Healthy Life Education, Kwame Nkrumah University of Science and Technology (KNUST) in Kumasi, Ashanti Region. Healthy Life Education, KNUST (2010) provided patient database records for the use of this study. The data contains 1027 people showing 577 females and 450 males between ages 19 to 80 in the Kumasi municipality and specify whether they have contracted hypertension in the past immediate years. It is assumed that the data represents a random and unbiased sample from the population defined.

3.6.1 Population

The targeted population comprises any potential high blood pressure person in Kumasi areas such as tertiary institutions, churches, individuals from different occupations in the Kumasi sub-metros in the Kumasi metropolitan of the Ashanti Region of Ghana by the Healthy Life Education, KNUST.

3.6.2 Anthropometry Measurement

This is the measurement system of the size and makeup of the body: height and weight ideal body weight body mass index, etc. Detailed description of how measurement indicates in real life has been clearly tabled in Table 3.2 and Table 3.3 for blood pressure reading and body mass index.

Table 3.2: The stages for hypertension.

Blood pressure reading (mmHg)	Description
<120/80	Normal
120/80 – 139/89	Pre-Hypertension (Mild)
140/90 – 159/99	Stage 1 hypertension
160/100	Stage 2 hypertension

Hypertension State Blood pressure parameters of individual are with systolic BP/ diastolic Bp of ≥140/90 mm Hg is considered hypertensive according to the World Health Organization (WHO) criteria.

Body Mass Index (BMI) was calculated using the formula;

$$BMI = \frac{Weight\ (Kg)}{Height(m^2)} \tag{3.21}$$

Table 3.3: Categorizes for Body Mass Index.

Body mass Index (BMI) Kg/m^2	Description
< 18	Underweight
18.1 – 24.9	Normal
25- 29.9	Overweight
>30	Obese

KNUST

3.7 Data Collection Procedure

The data is a set containing 1027 of observations. It contains patient records describing a number of attributes in addition to whether the patient went on to develop high blood pressure in some years. The data set contain the following variables in the table 3.4. Traceis Data exploration (software) was used to analyze the data.

Table 3.4: Definition of Indicators.

Variable	Description
Waist Circumference (WC)	Measurement around the waist
Systolic BP	Systolic Blood pressure
Diastolic BP	Diastolic Blood pressure
BMI	Body Mass Index
Age	Age of patient
Gender	Male (M) and Female (F)
Height	Height of patient
Weight	Weight of patient

W J SANE NO BADWENMA

Table 3.5 summarizes the variables in the data using set along with their anticipated role in the analysis using the categories described in above.

Table 3.5: Anthropometric Indicators modes of measurements

Variable	Continuous/ discrete	Scale of measurement	Anticipated role	comments
WC	Continuous	Ratio	Descriptor	Units : cm
Diastolic BP	Continuous	Ratio	Descriptor	Units : mm Hg
Systolic BP	Continuous	Ratio	Descriptor	Units : mm Hg
BMI	Continuous	Ratio	Descriptor	Body mass index
Age	Continuous	Ratio	Descriptor	Units : years
Gender	Discrete	Ordinal	Response	Male, female
Height	Continuous	Ratio	Descriptor	Units: meters
Weight	Continuous	Ratio	Descriptor	Units: kg

3.7.1 Transformation of variables

The data is further examined to determine whether any transformations of the data are required. The normalization transformation is considered within this analysis. This is to ensure that all variables are considered with equal weight for any further analysis. By using the min-max normalization with new range between 0 and1, this was applied to each variable. Table 3.6 illustrates a portion of the new table with the newly transformed variables added to the data set.

$$value' = \frac{value - OriginalMin}{OriginalMax - OriginalMin}(NewMax - NewMin) + NewMin \quad (3.22)$$

Table 3.6: Normalization of Variables

Age	Normalized Age	BMI	Normalized BMI	Systolic	Normalized Systolic	Diastolic	Normalized Diastolic
41	0.396	23	0.233	130	0.601	102	0.605
45	0.465	29	0.407	138	0.643	100	0.577
30	0.180	24	0.262	143	0.670	93	0.478
34	0.245	29	0.407	132	0.611	87	0.394
37	0.327	26	0.320	120	0.547	78	0.267

Summarization of sample observation grouped is shown in table 3.7 and grouping description can be found in table 3.2 and table 3.3 under anthropometry measurements.

Table 3.7: Binning of Indicators

Age	Grouped Age	Gender	BMI	Grouped BMI	Systolic BP	Grouped Systolic BP	Diastolic BP	Grouped Diastolic BP
52	40-59	F	29	Overweight	134	Mild	82	Mild
39	20-39	F	34	Obese	144	Stage 1Hypertension	86	Mild
30	20-39	F	32	Obese	148	Stage 1Hypertension	100	Stage 2 Hypertension
34	20-39	M	31	Obese	111	Normal	78	Normal
22	20-39	F	23	Normal	109	Normal	74	Normal

3.7.2 Variables Cleaning

Variable cleaning helps check for outliners. In this case the variable waist circumference is a candidate for removal because of most of its missing values. Table 3.8 shows the number of zeros values for each variable.

Table 3.8: Number of zero values

Variable	Number of zero values
BMI	8
Systolic BP	0
Diastolic BP	0
Waist circumference	83
Weight	8
Height	8
Age	0

KNUST

Correlation Analysis was done to identify relationships between BMI and Waist circumference and at the end we had a correlation coefficient *(r)* = *0.66* as shown in figure 3.1. This indicates a positive relationship meaning that one of these variables could be surrogate for the other; hence body mass index will be used for analysis in this research.

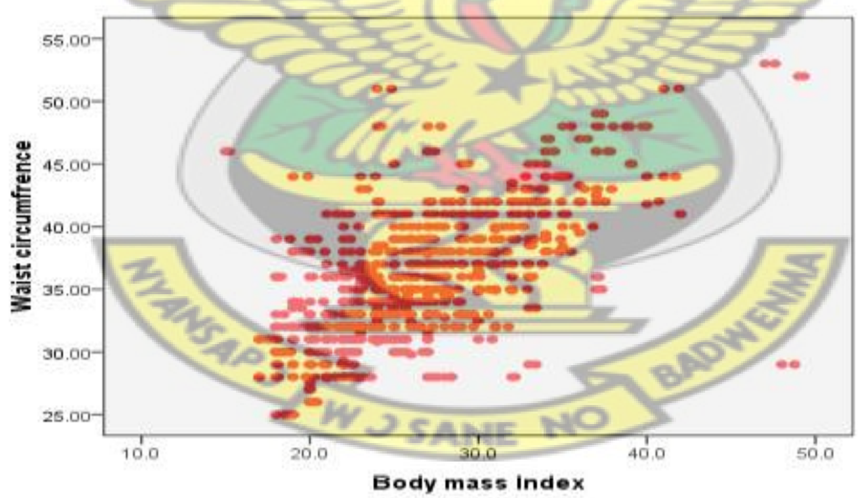


Figure 3.5: Relationship between BMI and Waist circumference

3.7.3 Variables for Analysis

The results at this stage are a cleaned and transformed data set ready for analysis. The variables Age, BMI, Systolic BP and Diastolic BP were cleaned and used for analysis in grouped and normalized variables with the state of hypertension as the categorized response for the data.

KNUST

CHAPTER FOUR

DATA ANALYSIS, DISCUSSIONS AND PRESENTATION OF RESULTS

4.1 Overview

This chapter is made up of the analysis of data from the Healthy Life Education, interpretation of results of the analysis and findings of the study in relation to the research questions. Based on the results of the analysis, the research questions were discussed to find out the relationships that existed between the various anthropometric indicators that is Age, gender, BMI, waist circumference, Systolic and Diastolic Blood Pressures.

4.2 Statistical Analysis of Data

The Means of the indicators were analyzed in this section as stipulated in the research question. The individual indicators have been clearly displayed in appendix A. The analysis and interpretations were made at 99% confidence interval because it covers wider area making it more accurate comparing it to 95% confidence interval in which most researchers used and using equal variance for each test.

4.3 Presentation of Results

In figure 4.1, the frequency distribution for those whose blood pressure indicates hypertension state according to Table 3.2 categorization of hypertension is shown. Figure 4.2 also shows frequency distribution of the four indicators that's BMI, Age, Systolic and diastolic blood pressures used for analysis in this research, indicating their central tendencies presented alongside a series of descriptive statistic in order to characterize the variable as shown in Table 4.1.

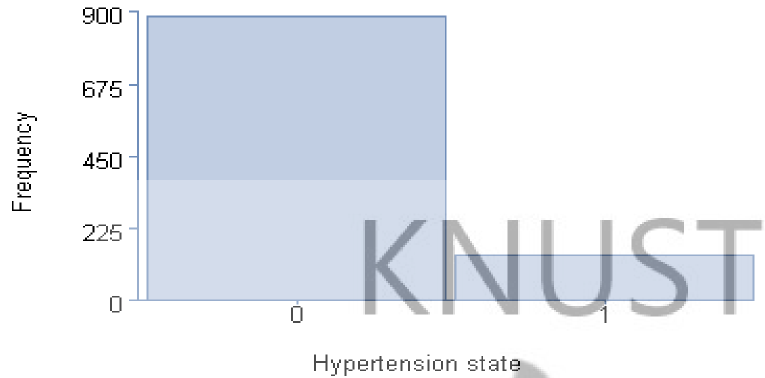
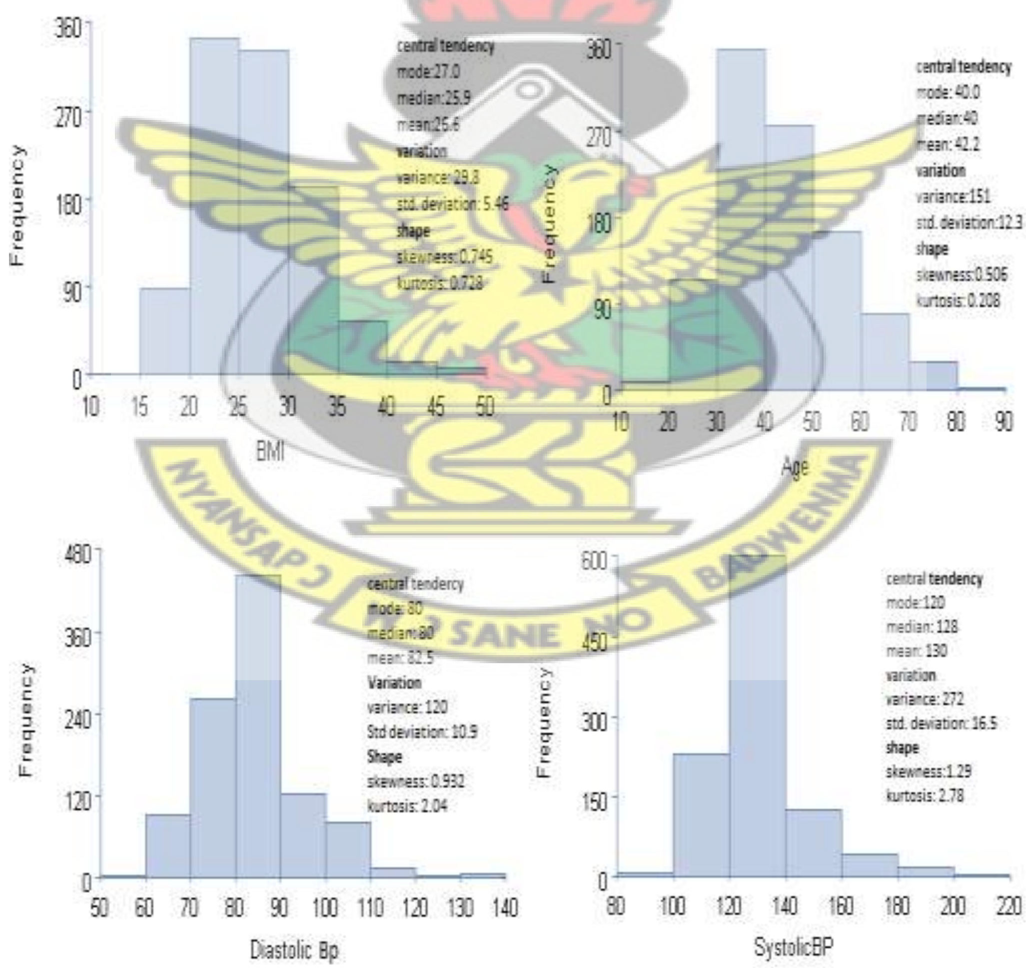


Figure 4.1: Frequency distribution of Hypertension State Category



KNUST

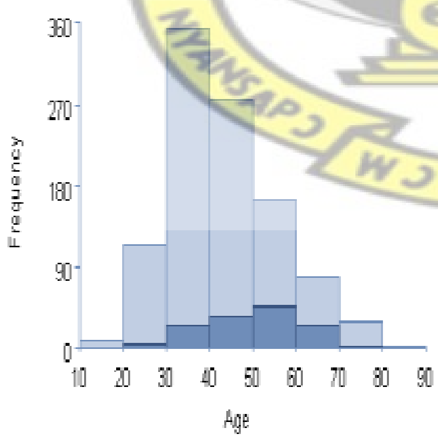
Figure 4.2: Descriptive statistics for descriptor variable

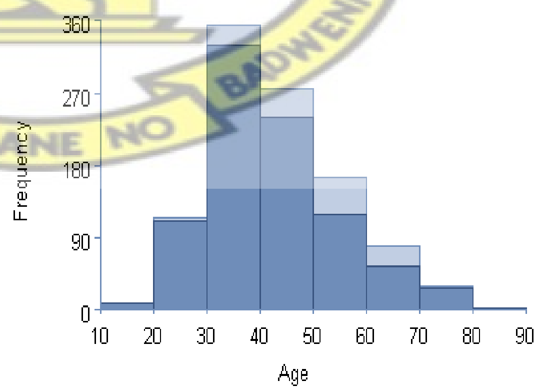
Table 4.1: Statistics for the anthropometric Indicators

Variable	Observation	Mean	Std. Dev.	Min	Max
Age	1027	42.2	12.3	19	80
BMI	1027	26.6	5.46	15	49.3
Systolic Bp	1027	130	16.5	97	205
Diastolic Bp	1027	82.5	10.9	59	130

Figure 4.3 shows frequency distribution for all descriptor variables across the two groups. In the histogram on the left, the dark blue highlighted belongs to the group that went on to develop hypertension. The observation highlighted in dark blue, on the right histogram are people who did not develop hypertension. And each of this histogram shows the total number of people according to their BMI, Age, Systolic and Diastolic Blood pressure who went on to develop hypertension. This helps in bringing out pictorially the ranges in which hypertension starts, increases at what BMI or common at what age.

Develops hypertension Does not develop hypertension





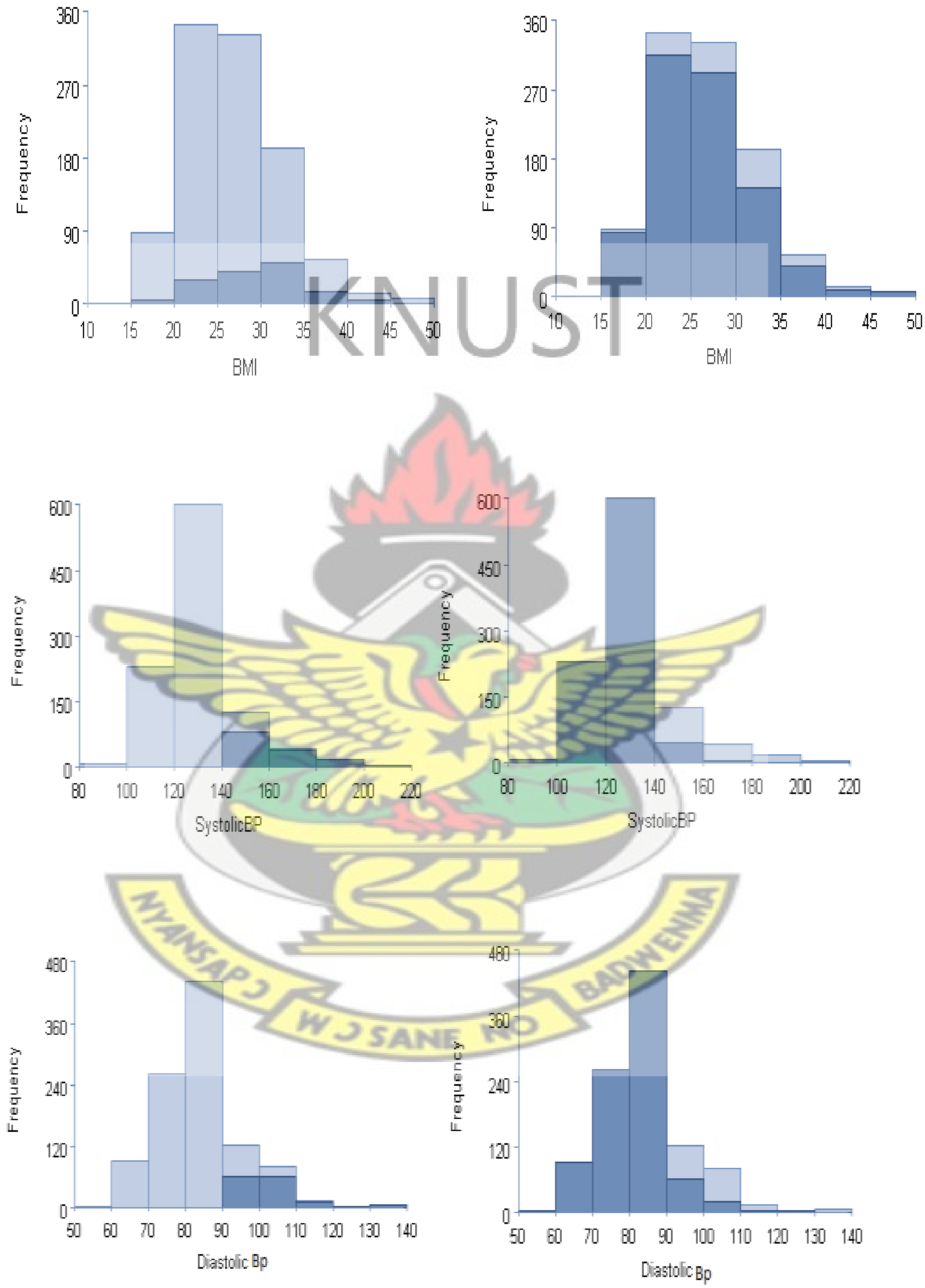


Figure 4.3: Frequency distribution across the two groups

Using the distribution for the variable Diastolic Bp and Systolic Bp from figure 4.3,the light blue colour is the overall frequency distribution and the dark blue highlighted observation on the left are the group that went on to develop hypertension and the dark blue highlighted group on the right did not develop hypertension. From these graphs it is difficult to see any discernible trends that differentiate the two groups, since the shape of the distribution is similar, even though the number of observation is higher in the group without hypertension. If we plot the two groups using a box plot, we see that the group that went on to develop hypertension is generally higher in the group that did not as shown in Figure 4.4

Figure 4.4: Box plot showing **Diastolic Bp** and **Systolic Bp** variations

Table 4.2 summarizes the means for all variables between the group that went on to develop hypertension and the group that did not. Figure 4.3 displays the frequency distribution for all variables to understand differences between the two groups (hypertensive and non hypertensive). It can be seen from the graphs, that the values for BMI, Age and are significantly different between the two groups whiles it is difficult to see the difference that of Diastolic Bp and Systolic Bp.

Table 4.2: Summary table of Means for Indicators of each group

Hypertension state	Patient count	Mean (Age)	Mean (BMI)	Mean (Systolic BP)	Mean (Diastolic BP)
No	886	41.124	26.137	124.96	79.802
yes	141	49	29.372	159.603	99.66

Graphs and summary tables have been used so far to visualize and characterized the difference between the two groups. In this research, we want to know looking at table 4.2 whether there is any difference at all in those who develop hypertension and those who did not from the 1027 observation. Then we want to understand if these differences are significant enough to make claims about the general population concerning their differences, hence we use z-test to make these assessment establishing differences between two means. The observations for the variables are divided into two groups, those people that went on to develop hypertension (group 1) and those people that did not go on to develop hypertension (group 2). We then specify a null and alternative hypothesis:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Where μ_1 is the population mean of group 1 and μ_2 is the population mean of group 2. At a 99% claim of confidence level ($\alpha = 0.01$) we calculate for the hypothesis score using each variable's sample means with respect to hypertensive and non hypertensive, the number of observation in group 1 and 2, the standard deviation and variance also for group 1 and 2. Table 4.3 shows the results of the z-test for all the variables for analysis with their means which will be used in the hypothesis testing, a *p-value* is determining whether to accept or reject the null hypothesis. In this case it can clearly be seen in Table

4.3 that, α is greater than *p-value* and we can say that our risk which is our *p-value* is low hence we reject the null hypothesis and we state that there is a difference.

Table 4.3 Hypothesis score for each variable using two means.

Variable	**Hypertensive state**	**Count**	**Mean**	**Standard Deviation**	**Hypothesis test (z)**	**p-value**
Age	Hypertensive	141	49	10.8	7.887	<0.0000003
	Not Hypertensive	886	41.124	12.2		
BMI	Hypertensive	141	29.372	5.3	6.722	<0.0000003
	Not Hypertensive	886	26.137	5.36		
Systolic BP	Hypertensive	141	159.603	15.6	25.502	<0.0000003
	Not Hypertensive	886	124.96	10.6		
Diastolic BP	Hypertensive	141	99.66	9.44	23.545	<0.0000003
	Not Hypertensive	886	79.802	8.39		

4.4 Modeling the Diseases

The other objective was to develop a predictive model to classify people into two categories:

1. Patient that will develop hypertension in the next five years.
2. Patient that will not develop hypertension in the next five years.

In order to achieve the third objective, that is for general prediction Artificial Neural Network is the ideal model to use to get the required results. The reason being that, neural networks are ideal in recognising diseases using scans since there is no need to provide a specific algorithm on how to identify the disease. Neural networks learn by example so the details of how to recognise the disease are not needed. The examples need to be selected very carefully if the system is to perform reliably and efficiently. For reliable and efficient results artificial neural network learn or train more slowly hence an experiment is conducted to train the learning process to achieve the best results displayed as sensitivity/ specificity as optimized in Table 4.4

4.4.1 Hypothesis Testing

The observations for the variables were divided into two groups,

• Those individuals that went on to develop hypertension (group 1)

• Those individuals that did not go on to develop hypertension. (group 2)

μ_1= the population mean of group1

μ_2= the population mean of group 2

$$H_0: \mu_1 = \mu_2$$

$$Ha: \mu_1 > \mu_2$$

We reject the null hypothesis hence we state there is a significant difference between the means, meaning there will be an increased in the hypertensive states. But we still need to know the number of increment. This is where I used Neural Network for predicting the number of people who will be hypertensive in this thesis; I therefore designed an experiment for the results predicted.

4.5 Designed Experiment of the ANN

An ANN was designed to analyze the features of each input in order to get a more accurate combination of inputs. The main idea of designing an ANN was to use the simplest structures to give the best results without over fitting the data. The designed ANN included: four inputs layer, two hidden layer and an output layer structure.

The steps for designing the ANN are shown below:

1. Collect the data – Extract enough useful features as the input of ANN.

2. Create a network – design and initialize the ANN.

3. Initialize the network – Initialize weights and biases.

4. Train the network – adjust the weights and biases.

5. Simulate the network – validate the network and apply new input data.

An experiment was performed to get the best result on sensitivity/specificity for the model which will aid in the prediction. This was conducted by varying the parameters (Age, BMI, Systolic and Diastolic Blood pressure) of the neural network. The varying of these parameters includes the combination of the four indicators, the combination of three indicators then from two to one indicator. By doing so helps in getting the best result as to whether all the indicators contributes to the risk factor for this disease and also train learning process to get the best results .The cycles used were 5000, 20000, 50000 and 100000. Hidden layers were varied between 1, 2 and 3 using the same learning rates of 0.5 and a cross-validation of 10% with the state of hypertension as the responses for the model. The input variables Age, BMI, Systolic and Diastolic BP were used to build the final model by adjusting them using fewer inputs to all of the inputs in creating an optimization table on sensitivity over specificity from the model and to select the best value for the experiment. The results of the experiment are then displayed in table 4.4. The results for 1 and 2 combinations of the input variable gave a sensitivity/ specificity of 0.000/1.000 in all cases hence there was no need of displaying them. This shows that a single input cannot be the only risk factor for the disease but all four combinations gave the best results. Hence these are the major contributors to the disease.

The following model gave the best overall performance and was selected: neural networks with two hidden layers, 50,000 cycles, and a learning rate of 0.5 using four descriptors as inputs. The overall concordance for this model from the experiment gave 0.975 (or 97.5%) with a specificity of 0.982 (or 98.2%) and a sensitivity of 0.929 (or 92.9%) as shown in Table 4.5.

Table 4.4: Results of measurements using ANN

KNUST

Input Variable				1 hidden Layer			
				5,000	20,000	50,000	100,000
SBp	DBp	BMI	AGE	0.000/1.000	0.234/0.993	0.0983/0.348	0.319/0.997
	DBp	BMI	AGE	0.000/1.000	0.000/1.000	0.000/1.000	0.000/1.000
SBp		BMI	AGE	0.000/1.000	0.000/1.000	0.000/1.000	0.056/0.998
SBp	DBp		AGE	0.000/1.000	0.000/1.000	0.000/1.000	0.000/1.000
SBp	DBp	BMI		0.000/1.000	0.000/1.000	0.000/1.000	0.000/1.000
Input variable				2 hidden Layers			
				5000	20,000	50,000	100,000
SBp	DBp	BMI	AGE	0.000/1.000	0.681/0.983	0.929/0.982	0.858/0.977
	DBP	BMI	AGE	0.000/1.000	0.355/0.976	0.220/0.985	0.596/0.959
SBp		BMI	AGE	0.000/1.000	0.255/0.993	0.624/0.981	0.723/0.976
SBp	DBp		AGE	0.000/1.000	0.539/0.988	0.674/0.820	0.447/0.998
SBp	DBp	BMI		0.000/1.000	0.596/0.997	0.461/0.992	0.362/1.000
Input Variable				3 hidden Layers			
				5,000	20,000	50,000	100000
SBp	DBp	BMI	AGE	0.000/1.000	0.319/0.995	0.823/0.990	0.894/0.993
	DBp	BMI	AGE	0.000/1.000	0.063/0.995	0.248/0.981	0.489/0.962
SBP		BMI	AGE	0.000/1.000	0.000/1.000	0.574/0.981	0.645/0.979
SBp	DBp		AGE	0.000/1.000	0.049/0.998	0.489/0.974	0.723/0.991
SBp	DBp	BMI		0.000/1.000	0.000/1.000	0.567/0.984	0.879/0.984

4.6 General findings

Taken a critical look at our model built, Figure 4.5, we can see that some observations were not correctly predicted. That is observations predicted to be in the not hypertensive group but who were hypertensive (false positive) was ten (10)**.** Figure 4.6 presents a series of box plots for observations predicted to be in the not hypertensive group but who were hypertensive. The upper box plot represents the set of false negatives, the lower present all observation.

KNUST

		Actual (Hypertension state)		
		1	0	Totals
Predicted (Hypertension state)	1	131	16	147
	0	10	870	880
	Totals	141	886	1027

Figure 4.5 Overall best model contingency table.

Neural Network Model Sumary	
Independent variables	Age, BMI, SystolicBP, Diastolic
Response	Hypertension state
Number of cycles	50000
Number of layers	2
Learning rate	0.5

Cross Validated Results	
Accuracy	0.975
Error	0.0253
Sensitivity	0.929
Specificity	0.982
False positive rate	0.0181
Positive predictive value	0.891
Negative predictive value	0.989
False discovery rate	0.109

Table 4.5 ANN Best Model Summary

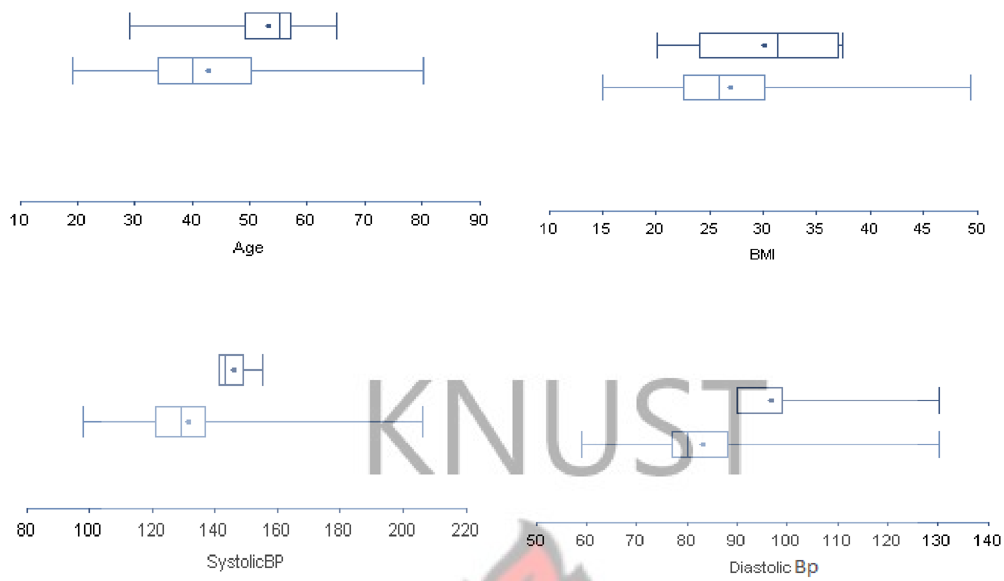


Figure 4.6 Summary of contents of false positives

Based on our understanding of the data, hypertension is often associated with increase in high blood pressure and a high rate of body mass index (contributing to other risk factors such as level of physical activity, cholesterol level, etc). Also where we predicted the person to become hypertensive when in fact they did not (false negative) were sixteen (16). Figure 4.6 presents a series of box plots for the descriptor variable, the upper box plots are the false negatives and the lower box are all observations. These people have characteristics, based on our understanding of the data, of individual that would go on to develop hypertension, which is elevated high blood pressure and increased BMI. Hence this will suggest that the data is missing important fields for the classification of these group individuals.

KNUST

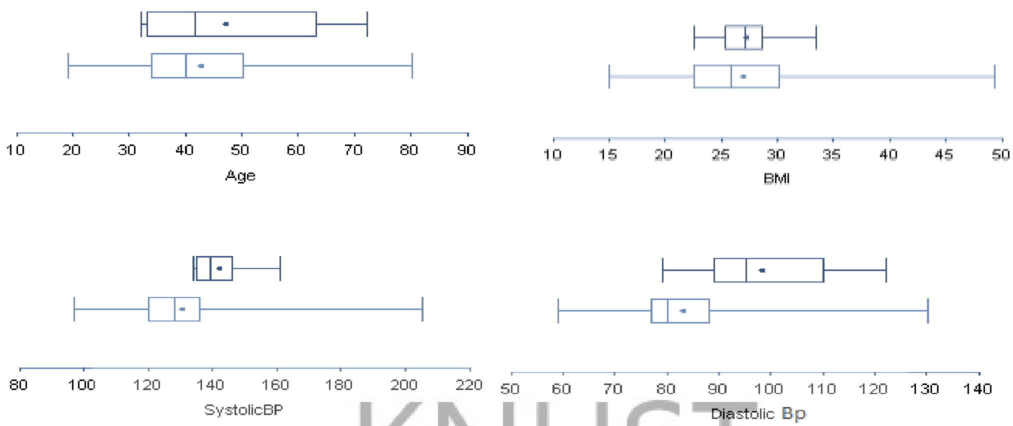


Figure 4.7 Summary of contents of false negatives

KNUST

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Overview

This chapter consists of the summaries and conclusions of the study. Recommendations and suggestions are made accordingly for further research.

KNUST

5.2 Summary

The study was conducted to investigate the relationship between people who have and do not have hypertension and those who will get in the next years. The data was the recording of any potential adult blood pressure and other anthropometries indicators by the Healthy Life Education in the Kumasi Metropolitan from some selected Sub metro in 2010. The reading recordings of body mass index (BMI), patient's age, weight, height, waist circumference, and systolic Blood pressure and Diastolic Blood pressure were used. A total number of five hundred and seventy seven (577) female and four hundred and fifty (450) males making a total of one thousand and twenty seven (1027) people.

The mean values and standard deviations for the data were calculated using Traceis Data exploration. Neural Networks was use for the building of the model in which after we optimized for the beat value in Table 4.4. The mean results for age according to their hypertension state are shown in Table 4.3. The total number of hypertensive persons were 13.7% and those who do not have hypertension is 86.3%: Pre-hypertension 49.9% and Normal 40.4%. Hence is evidential that the hypertensive state of the people selected in the various sub metro of the Kumasi metropolitan is significantly better comparing percentages using the criteria for hypertension state from the WHO, which is 140/90 mm Hg. Whereas 45.9% were considered pre-hypertensive; again out of 577 females as compared to 450 males, 99 (70.2%) had hypertension where 42 (29.8%) were hypertensive males. For the 886 non-hypertensive patient, there were 408 males (46.1%) and 478 females (53.96%). The hypothesis made using z-test shows that with an α greater than p-value calculated, it was obvious that there was a significant difference comparing the two means for those who went on to develop hypertension and those who did not go on to develop hypertension. Hence a general prediction using the modeling can be made in making a general conclusion in the prevalence of hypertension in the Kumasi metropolitan.

This research examines the hypertension states of people living in some selected area of the Kumasi metropolitan in Kumasi, Ashanti Region, West Africa. The hypertension state from the high blood pressure reading had a positive strong linear relationship. The high records of high blood pressure may be due the following reasons:

i. White coat hypertension, when a person's blood pressure is normal when taken at home but is high when taken by the doctor (who may be wearing a white coat). This phenomenon is thought to be due to stress of seeing the doctor.

ii. Exercise hypertension, when there is excessive high elevation in blood pressure during exercise.

iii. Pregnancy -induced hypertension, when the patient is pregnant, her blood pressure may increase and in the case of the pregnant women was not stated.

Again, reasons for low numbers for hypertension state as compared to non hypertension states is that ,the criteria of grouping hypertension was according to the WHO where systolic is 140 and diastolic is 90 hence a patient with either one of them high or the other low is classified non hypertensive.

5.3 Conclusions

The following conclusions were made from the analyses, discussions and interpretations of the results. The analysis of the mean and standard deviations of the results, according to their hypertension state indicates that:

1. The age group for hypertension is below 50 years with an overweight BMI and pre- hypertensive, contributing to sudden cardiovascular attacks.
2. It was predicted that, 1.85% patient will be hypertensive, which increase the state of hypertension.
3. Artificial Neural networks can be used efficiently to model predictive hypertension cases to make more consistent diagnosis.

It was observed during the experiment that the combination of all indicators gave the overall best model of 0.929/0.982 of sensitivity/ specificity hence they all contributes to the risk factor. From our hypothesis we were asked to reject the null hypothesis hence we state that there is a difference. So if there is a difference, then we can move on developing our prediction using the Neural Network Mode to develop a predictive mathematical model to estimate whether a patient will develop hypertension from normal through pre-hypertension. In our model, it indicated that 16 people out of the 886 non-hypertensive persons from normal through mild will develop hypertension, all things being equal that is living the same life style or it can improve for the better with a healthy and positive life style change.

5.4 Recommendations

These are the following recommendations the researcher made based on the conclusions.

1 White coat hypertension is one of the major factors for elevation of blood pressure and as such, it should be prevented when at all cost when blood pressures are being recorded.

2 Pregnancy -induced hypertension: for other implementation of data purposes and what is being use for as in diagnosis sake, females whom are pregnant should be indicate so that the researcher will make the appropriate analysis and conclusion.

3 Individuals should monitor their blood pressure, Body mass index, exercise and eat healthy balanced diet to enjoy healthy life.

4 Artificial Neural networks techniques can be used efficiently to model predict hypertension cases, hence should be encouraged.

5 This study can be used as an assistant tool by cardiologists to help them make diagnosis for hypertension.

6 Public Health education: Health workers should also make it an importance to keep the record of patient's blood pressure readings and make it a point to explain to them their hypertension status whether they are elevating or not and ways to maintain a normal blood pressure.

5.5 Further study

This research was limited to only some selected sub-metro of the Kumasi metropolitan, Kumasi in the Ashanti Region of Ghana West Africa. In other to achieve good results and profile of performance, the study should include a data of high blood pressure readings of the patient for about three months to reduce false readings and also check hypertension states since most patient where not able to provide their hypertension status and that of their family history compared to what was used for the studies. This research is conducted in health studies using a mathematical model; similar studies can be done to capture other health issues or diseases within Kumasi and other regions in Ghana. The sample is small and does not represent the entire population; it is therefore recommended that the survey and other designs should be conducted with extra care to cover a very large population if not all sub-metros in the metropolitan for generalization.

In summary, a finding from the main objective in this overview indicates healthier lifestyles and effective treatment should help prevent and control hypertension. Hypertension is clearly an important public health problem in Ghana, even in the poorest rural communities. Emerging opportunities such as the national health insurance scheme and public health education emphasizing on health promotion will aid in controlling this health issue. It is recommended that the research should be conducted to find hypertension states using patient's blood groups and occupation and their correlations to elevated high blood pressures.

REFERENCE

1. Abdoljalal, M. (2011). *Waist circumference, body mass index, hip circumference and waist-to hip ratio in type 2 diabetes patients in Gorgan, Iran.* Journal of Clinical and diagnostic Research.2011 April, Vol-5(2):201-205
2. Addo, J. (2006). *The changing patterns of hypertension in Ghana: A study of four rural communities in the Ga district.* Available from: http//mxi.ishib.org/ED/journal/16-4/ethn-16-04-894.pdf. [accessed on May 2012]
3. Anupama, K; Nandini M (2005). *Relation of anthropometry to CVD risk factors in young obese women.* Biomedical Research 2005;16(2):137-141.online available from :http//www.Pubmed health.com
4. American Heart Association (2011). *Hypertension; Limitation of Analyses based on achieved Blood pressure,* available from: http://hyper.ahajournals.org/cgi/content/full/57/6/1061
5. Association, A. H. (2011). *Autonomic-Immune-Vascular Interaction,An Emerging Concept for Neurogenic Hypertension*. Retrieved June 8, 2011, from http://hyper.ahajournals.org
6. Bambrick, H J. (2005). *Relationships between BMI, Waist circumference, hypertension and fasting glucose: Rethinking risk factors in Indigenous diabetes.* available online from: http://www.healthinfonet.ecu.edu.au/html

7. Beune E J. (2006). *Under pressure: how Ghanaian, African-Surinamese and Dutch patients explain hypertension.* Journal of human Hypertension (2006). Available from: http//www.nature.com/jhh/journal/v20/n12/full/1002094a.html

8. Bektas M Y, Erkan M S, Esra Y (2005). *Which anthropometric measurement is most closely related to elevated blood pressure?* Available online from: http//www.oxfordjournals.org/doi: 10.1093/fampra/cmi043. [accessed on July 2011]

9. Bonsu, W K (2010). *Epidemic of hypertension in Ghana: a systematic review.* Bonsu BMC Public Health 2010 online available from: http//www.biomedentral.com/1471-2458/10/418.[accessed on August 2011]

10. Books, M. H. (1990). *The Bantam Medical Dictionary.* New York: Bantam Books Ltd.

11. Chen-Huan, C, (0200). *Different association of hypertension and insulin-related metabolic syndrome between men and women in 8437 non-diabetic Chinese.* American Journal of Hypertension 13, 846-853 (July 2000)/doi:10.1016/s0895-7061 (00)00255-7. [accessed July 2011]

12. Cheung. V. and Cannons .K (2002). *Introduction to Neural Networks.* University of Manitoba, Manitoba, Canada.

13. Cook. R (2007) .*Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction.* Circulation Journal , published by the American Heart Association, *Available online from:* http://circ.ahajournals.org/content/115/7/928

KNUST

14. Damtew .A (2011). *Designing a Predictive Model for Heart disease Detection Using Data Mining techniques.* A published postgraduate thesis, School of graduate studies, Addis Ababa University.

15. Esmaillzadeh, A (2004). *Waist-to-hip ratio is better screening measurement for cardiovascular risk factors than other anthropometric indicators in Tehranian adult men*.Int J Obes Relat Metab Disord. 2004 Oct; 28(10):1325-32. Available online from: http//www.pubmed.com. [accessed on June 2011]

16. Francesco P.C (2008). *Body* size blood pressure: *An analysis of Africans and African Diaspora.* UKPMC Funders Group (2010). Epidemiology: 2008 January; 19 (1):38-46. available online from: http//www.epidem.com

17. Ghana Health Service (2007). *Hypertension is the number one killer disease in Ghana.* Available from: http//www.ghanahealthservice.or/article.php.[accesed on May 2012]

18. GNA (2011). *Hypertension, diabetes on the upsurge in Kumasi.* Available from: http//www.ghanabusinessnew.com.[accessed on May 2012]

19. Henkle, B (2009). *University of Iowa studies hypertension in Ghana.* Available online from: http//ihdn.org/Iowastudent09.pdf. [accessed on May 2012]

20. Ikeme C(1978). *Cardiovascular status and blood pressure in a population sample in Ghana--the Mamprobi survey. Trop Geogr Med* 1978, **30:**313-329. PubMed Abstract

21. Jasenka, Z; Hidefumi, W (2011). *Autonomic-Immune-Vascular Interaction: an emerging concept for Neurogenic Hypertension.* Available online from: http//hyper.ahajournals.org.(accessed August 2011)

22. JSL de Munter; C Agyemang; I G van Valkengoed; R Bhopal; k Stronks(2011). *Sex difference in blood* pressure *among South Asian Diaspora in Europe and North America and the role of BMI: a meta-analysis.* Journal of human Hypertension 25,407(2011). doi:10.1038/jhh.2010.77. available from: http//www.nature.com/jhh

23. Krose .B and Van der Smagt. P (1996). *An Introduction to Neural Networks.* 8th Edition. Amsterdam, Netherlands.

24. Kunutsor, S; Powles J (2009). *Descriptive epidemiology of blood pressure in a rural adult population in Northern Ghana.* Rural Remote health, 2009 Apr-Jun; 9(2):1095.Epub 2009 June. available online from: http//www.pubMed.com.[accessed on May 2011]

25. Lichtenstein M J; Shipley M J, Rose G(1985).*Systolic and diastolic blood pressures as predictors of coronary heart disease mortality in the Whitehall study.* British Medical Journal, Volume 291, 27 July 1985 .[accessed on May 2011]

26. Margen, S. (1991). *The Wellness Encyclopedia.* Boston: Houghton Mifflin Company.

27. McOscar, D. L. (1996). *Pharmacological Aspects of Nursing Care* (Fifth Edition ed.). Boston, Boston, USA: Delmar Publishers.

28. Mensah, G A (2008). *Epidemiology of stroke and high blood pressure in Africa.* Available from: http://heart.bmj.com/content/94/6/697.full.pdf+html.

29. Mostafa A. (2011). *Prevalence of elevated blood pressure and association with obesity in Egypt school adolescent*. International Journal of Hypertension. doi:10.4061/2011/952537

30. Myatt, G .J (2007). *Making Sense of Data: A practical guide to exploration and data mining*. Available online from: http//www.wiley.com.

31. Myjoyonline (2007). *Hypertension*. Available from: http//www.myjoyonline.com/health/200708/7461.asp. [accessed on May 2011]

32. Ofuya, Z M (2007). *The Incidence of hypertension among a selected population of adults in the Niger Delta region of Nigeria*. Available from: http://www.tm.mahidol.ac.th/seameo/2007_38_5/24-4090.pdf. [accessed on May 2012]

33. Ohno-Machado .L (1996). *Medical applications of artificial neural networks: Connectionist models of survival*. A Published Dissertation, School of graduates Studies, Stanford University.

34. Oleckno, W. A. (2002). *Essential Epidemiology, principles and Application.* Illinois: Waveland press.

35. Owiredu W. (2010). *Seasonal variation in the occurrence of pregnancy-induced hypertension- A Ghanaian Study*. Available from: http://www.ajol.info/index.aphp/jgsa/article/view/56811.

36. Pobee J. O (1993). *Community-based high blood pressure programs in sub-Saharan Africa*. *Ethn Dis* 1993, **3**(Suppl)**:**S38-45. PubMed Abstract

37. PubMed Health (2010). *Hypertension, High Blood Pressure*,(online) available from: http//www.ncbi.nlm.nih.gov/…/PMH0001502. (Accessed August 2011)

38. Rojas. R (1996). *Neural Networks* .Springer-Verlag, Berlin

39. Rowan, R. L. (1986). *How to Control High Blood Pressure.* NewYork: Ballantine Books.

40. Somerville, R. (1996). *The Medical Advisor.* Virginia: Time-Life Books.

41. Stergiou. C and Siganos. D. (1996). *Neural Network.* Available from: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

42. Tegner J N. (2009). *Computational disease modeling- fact or fiction.*BMC systems Biology 2009. Available from: http//www.biomedcentral.com/1752-0509/3/56.[accessed on September 2011].

43. Tin L L; Beevers D G; Lip H. (2002). *Systolic vs Diastolic blood pressure and burden of hypertension.* Journal of Human hypertension (2002) 16, 147-150. Available online from: http//www.nature.com/jhh. [accessed on February 2012]

44. Twagirumukiza, Marc (2011). *Current and projected prevalence of arterial hypertension in sub-Saharan Africa by sex, age and habitat: an estimate from population studies.* Journal of hypertension. 29(7):1243-1252, July 2011. [accessed on May 2011]

45. University, T. J. (1999). *Johns Hopkins Family Health Book.* New York: HarperCollins publishers.

46. Whitworth J. A; 2003 World Health Organization (WHO)/ International Society of Hypertension (ISH) Writing Group (2003). *2003 WHO/ISH on Management of hypertension.* Available online from: http// www.ncbi.nlm.nih.gov [accessed on June 2012]. WWW.WHO.ORG. (Accessed 2011 July)

KNUST

APPENDIX A

grouped BMI

1\.

Gender

	Normal	overweight	obese	underweight	Totals
M	231	163	52	4	450
F	190	168	213	6	577
Totals	421	331	265	10	1027

KNUST

grouped Systolic BP

2\.

Gender

	Mild	Stage 1 hypertension	Normal	Stage 2 hypertension	Totals
M	281	32	112	25	450
F	319	93	124	41	577
Totals	600	125	236	66	1027

Hypertension state

3\.

Gender

	0	1	Totals
M	408	42	450
F	478	99	577
Totals	886	141	1027

4.

Grouped Diastolic BP

		Stage 2 hypertension	Stage 1 hypertension	Mild	Normal	Totals
Gender	M	54	38	160	198	450
	F	52	84	282	159	577
	Totals	106	122	442	357	1027

5.

grouped Age

		40- 59	20- 39	60- 79	80- 100	0- 19	Totals
Gender	M	167	234	46	0	3	450
	F	272	234	64	2	5	577
	Totals	439	468	110	2	8	1027

KNUST