RAINFALL PREDICTION FOR MINNA METROPOLIS USING ARTIFICIAL NEURAL NETWORK

BY

MOHAMMED NDACHE USMAN
MTECH/SPS/2017/7005

THE DEPARTMENT OF PHYSICS
FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA

OCTOBER, 2021

RAINFALL PREDICTION FOR MINNA METROPOLIS USING ARTIFICIAL NEURAL NETWORK

BY

MOHAMMED NDACHE USMAN
MTECH/SPS/2017/7005

A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF TECHNOLOGY IN APPLIED ATMOSPHERIC PHYSICS

OCTOBER, 2021

ABSTRACT

The effect of rainfall in our society today is stupendous. Rainfall is seen as a benefit to crops and lives. Accurate and timely rainfall prediction can be very helpful for effective security measures for planning water resources management, transportation activities, agricultural tasks, managing flights operations, issuance of flood warning and flood situation. This study aims to predict the rainfall of Minna metropolis. Atmospheric data comprising those of maximum temperature, minimum temperature, relative humidity and rainfall for four consecutive years spanning from January 2015 - December 2018 were acquired from the Geography Department of Federal University of Technology, Minna. The datasets were preprocessed and normalised, and further partition into three parts: 70% for training set, 15% for testing set and 15% for validating set. Feed forward neural network and binary classification was used for the prediction. The target data (rainfall) was labelled as positive or negative (rainfall or no rainfall), that is, (1 or 0) with threshold of 0.5 for classifying the rainfalls. The outcomes of prediction were evaluated using confusion matrix. The best test result indicates that 66 days were predicted to have rainfall and 120 days predicted for no rainfall with 69% accuracy, 1.3% error, 63% sensitivity and 84% specificity. The best validated results also indicate that 77 days were predicted to have rainfall and 109 days predicted for no rainfall with 59% accuracy, 1.4% error, 52% sensitivity and 78% specificity. The performance of the classifier is 0.568 (AUC = 57%).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

1.0 INTRODUTION

1.1 Background to the Study

Rainfall is a natural phenomenon whose prediction is challenging and demanding as the world continues to witness an ever changing climate conditions. Its forecast plays an important role in water resource management and therefore, it is of particular relevance to the agricultural sector, which contributes significantly to the economy of any nation Abdulkadir *et al*. (2012). Rain in Nigeria increases from the coastal region, with annual rainfall greater than 3500 mm, to the Sahel region in the north-western and north-eastern parts, with annual rainfall less than 600 mm (Omonona and Akintunde, 2009). The inter-annual variability of rainfall, particularly in the northern parts often results in climate hazards, especially floods and erosion with their devastating effects on farm products and associated calamities and sufferings.

Several neighbourhoods in Minna, the capital of Niger state in the north central of Nigeria are under the threat of flash flood from heavy rainfall. The residents of Dusten Kura area in Minna woke up to a big shock of rain water flooding their homes after a heavy rainfall on 13th July 2017. The whole incident affected the movement of vehicles and people due to water overflowing the drainage canal. On 2nd September 2012, two children were drowned while trying to cross a flooded drain caused by heavy rainfall in Minna (NSEMA, 2012). According to Niger State Emergency Management Agency (NSEMA), at least fourteen persons died due to flooding in different parts of Minna and over sixty houses were affected by flood due to heavy rainfall in Kontagora, Tafa and Suleja Local Government Areas of the State (NSEMA, 2018). On 9th September 2018, according to The Guardian Newspaper, over thirty villages were affected by flood in Mokwa Local Government Area of the State. Houses, farmlands and live stocks were destroyed in this disaster (Babalola, 2014). In most cases, floods were associated with abnormally high daily rainfall events (Umar, 2012).

The effect of rainfall on human civilisation is colossal. Rainfall means crops; and crop means life. Additionally, rainfall has a strong influence on the operation of dams and reservoirs, sewage systems, traffic and other human activities. Previous studies have shown that among the entire climate elements, rainfall is the most variable element in Nigeria both temporally and spatially which can have significant impact on economic activities (Kowal and Kanabe, 1972; Kowal and Kassam, 1978). Rainfall is one of the challenging tasks in weather forecasting. Weather data consists of various atmospheric features such as wind, precipitation, humidity, pressure, and temperature among others. Accurate and timely rainfall prediction can be very helpful for effective security measures for planning water resources management, issuance of early flood warning, construction activities, transportation activities, agricultural tasks, managing the flight operations and flood situation. Data mining techniques can effectively predict rainfall by extracting the hidden patterns among available features of past weather data (Aftab and Ahmad, 2018). The variability of rainfall is a crucial phenomenon in today's world. It is ever challenging and a topic of interest because prediction is not always accurate. It is a continuous, high dimensional, dynamic and complicated process because it involves many factors of the atmosphere. The parameters required to predict the weather are enormously complex such that there is uncertainty in prediction even for a short period (Geetha and Nasira, 2014).

In Nigeria and on the worldwide scale, large numbers of attempts have been made by different researchers to predict rainfall accurately using various techniques, but due to the nonlinear nature of rainfall, prediction accuracy obtained by these techniques is still below the satisfactory level. Of course, as with anything else, too much rain can lead to a host of problems. Heavy rainfall can lead to numerous hazards, for example: flooding, including risk to human life, loss of crops and livestock, landslides which can threaten human life, disrupt transport and communications, and cause damage to building and infrastructure. The increase in rainfall will also improve water availability, a condition which will impact on water supply and improve sanitation and health care delivery (Ifabiyi and Ashaolu, 2013). Therefore, it is important to evaluate how rainfall varies and how it will be in the future to minimise and reduce the negative impact of heavy rainfall and to increase the society resilience to hazards such as floods and erosions. To achieve this, researchers are developing and applying improved weather prediction models capable of accurately forecasting several events in Nigeria.

Artificial Neural Network algorithm becomes an attractive inductive approach in rainfall prediction owing to the non-linearity, flexibility and data learning in building the models without any prior knowledge about catchment behaviours and flow processes. In machine learning, classification can be referred to as task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain. Machine learning is a field of study and is concerned with algorithms that learn from examples. Classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. From a modeling perspective, classification requires a training dataset with many examples of input and output from which to learn. A model will use the training dataset and calculate how to best map examples of input data to specific class labels. Data mining algorithms are classified as supervised and un-supervised. Supervised methods get trained first with pre-classified data (training data) and then classify the input data (test data) (Ahmad and Aftab, 2017). Un-supervised methods on the other hand do not require any training; instead of pre-classified data, these techniques use algorithms to extract hidden structure from unlabeled data. It has been observed from latest research that for high accuracy, researchers prefer the integrated techniques for the rainfall prediction. In general, climate and rainfall are highly non-linear and complicated phenomena, which require advanced computer modeling and simulation for their accurate prediction. An Artificial Neural Network (ANN) can be used to predict the behaviour of such nonlinear systems (Nayak, 2013).

1.2 Statement of the Research Problem

Heavy rainfall can lead to numerous destructions, for example; flooding, landslides and erosion which can threaten human life. It can also damage buildings and infrastructure, disrupt transportation and communications and cause losses to farm crops of the affected areas. Heavy rainfall has caused lots of damages and destruction to lives and properties in some parts of Minna, Niger State capital.

According to British Broadcasting Corporation on 27th September 2018, NEMA declared state of emergency in four states (Niger, Kogi, Anambra and Delta) due to destruction by heavy rainfall (BBC New, 2018). Excess rain brings other negative effects on the environment and even the economy of the affected location. Existing rain prediction are mostly numerical and traditional in nature and are not accurate (Gugulethu, 2013).

1.3 Aim and objectives of the study

The aim of this study is to predict the rainfall of Minna metropolis using Artificial Neural Network. The aim shall be achieved through the following objectives which are to:

i. carry out interpolation of missing data and min-max normalisation technique for data scaling; and

ii. predict the number of rainfall days using classification method of Artificial Neural Network as a predictive tool.

1.4 Study Area

Minna is the headquarters of Chanchaga Local Government Area of Niger State, Nigeria. It is the capital city of the state. It lies between Latitudes 09°40′ 7.63″ N and 09° 39′ 59.72″ N and Longitudes 06° 30′ 0.32″ E and 06° 36′ 34.05″ E. Figure 1.1 (a) and (b) is the map of the study area.

Minna lies on a valley bed (that is, lowland) bordered to the east by Paida hill stretching eastwards towards Maitumbi and bordered by Wushishi and Gbako to the West, Shiroro to the North, Paikoro to the East and Katcha to the South.

Minna possesses the tropical continental wet and dry climate based on the Koppen Classification Scheme and is characterised with two distinct seasons namely; the wet season which begins around March and runs through October and dry season which begins from October to March. The city has a mean annual rainfall of 1334 mm with September recording the highest rain of close to 330 mm on the average, while the least amount of rainfall occurs in December and January which can be as low as 1mm. Minna is about 150 km away from Abuja, the capital of Nigeria (Source: Geography department, FUT Minna).

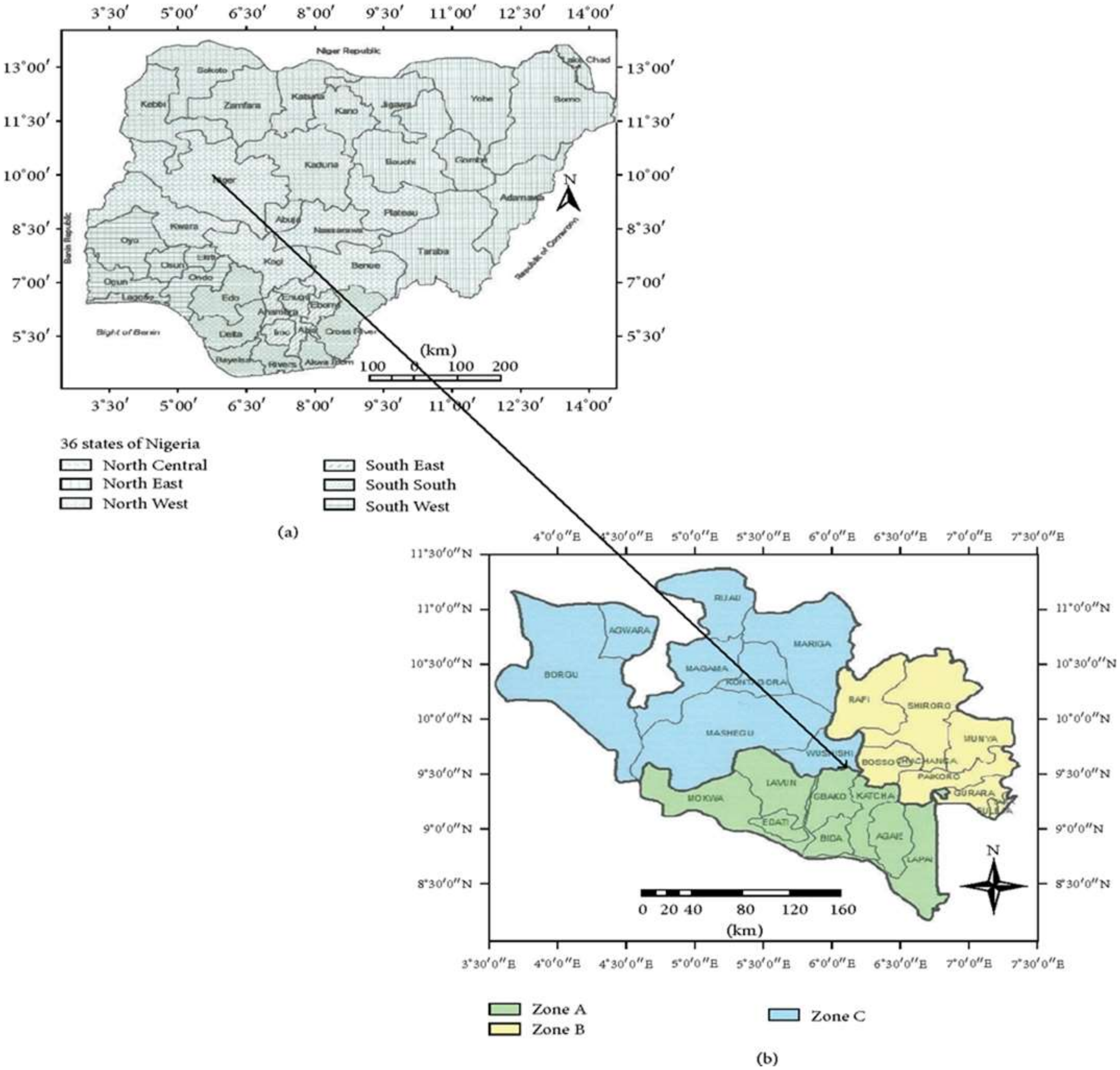

Figure 1.1 Map of Nigeria (a) and Niger State (b) showing the study area.

(Source: Geography department, F.U.T Minna)

1.5 Significance of the Study

A research of this nature is very important particularly to Minna residents, Niger State Government authorities and the research community as it will enhance the safety of lives and properties from rainfall hazards due to better awareness, preparedness and planning by farmers, aviation sector, construction firms and disaster managers. Also, it will help to make rainfall prediction data available to all stakeholders.

1.6 Justification of the Study

The change in rainfall has implications in various sectors of the economy of Niger state. There is an increase in decade anomaly of rainfall in Minna (Akinsanola and Ogunjobi, 2014). According to Daramola *et al.* (2017), there are more wet years in the South and middle Belt of Nigeria which are prone to the occurrence of flooding.

On the 25th to 26th August, 2014, heavy downpour spoiled most parts of Minna, the Niger State capital, causing serious damages. It was gathered that houses, fences, mini bridges were washed away by the heavy rain. Some of the affected areas were Barikin Sale and Farm centre in Tunga. Others are Niteco, Nykangbe and Kpankungu areas (Babalola, 2014).

The areas that are previously affected by heavy rains and are still prone to flooding in Minna metropolis are Fadikpe, Barikin Sale, Shango. These areas are further shown in Figure 1.2

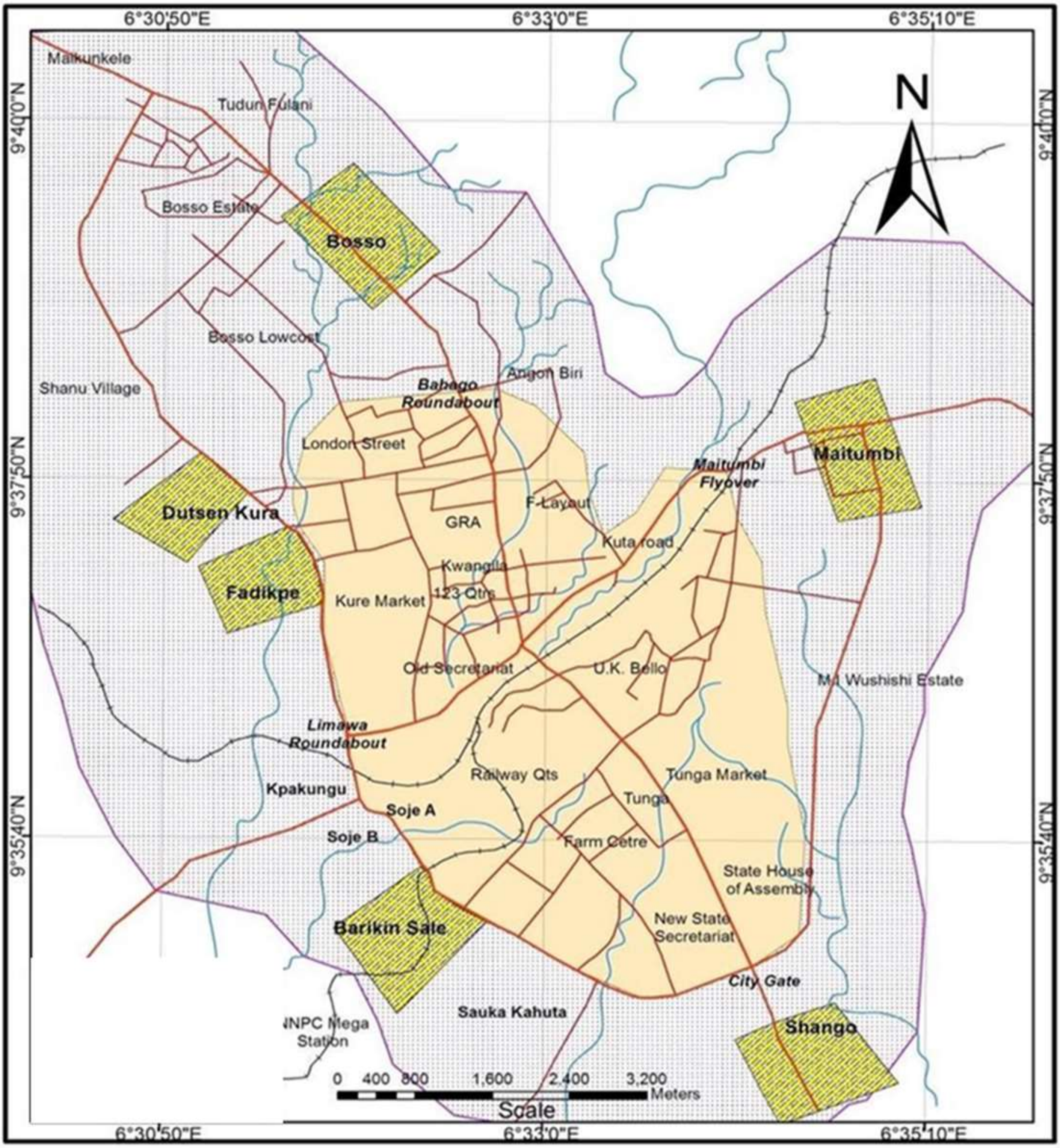

Figure 1.2 Map of Minna showing areas affected with flooding

(Source: Geography department, F.U.T Minna)

Hence, this study is necessary based on the flooding history of Minna. A prediction of heavy or low rainfall serves as an alarm to individual, communities and relevant government agencies.

1.7 Scope and Limitation of the Study

This research studied the rainfall variations and classification method was applied to predict number of rainfall and no rainfall days in Minna using four-year dataset obtained for maximum temperature, minimum temperature, relative humidity and rainfall spanning from January 2015 - December 2018.

However, this study will be limited to rainfall prediction using four atmospheric parameters.

CHAPTER TWO

2.0 LITERATURE REVIEW

This chapter deals with literature review. Researchers have been working to improve the accuracy of rainfall prediction by optimising and integrating different techniques. Some of the studies are discussed in this chapter.

2.1 Weather and its Parameters

One of the first things usually observe every morning is what the atmosphere is like. The weather affects us in many ways. Day-to-day changes in weather can influence how we feel and the way we look at the world. Severe weathers, such as: tornadoes, hurricanes among others, can disrupt people's activities because of the havoc they might cause. The term "weather" refers to the temporary conditions of the atmosphere, the layer of air that surrounds the Earth. We usually think of weather in terms of the state of the atmosphere in our own part of the world. But weather works like dropping a pebble in water, the ripples eventually affect water far away from where the pebble was dropped. Weather doesn't just stay in one place. It moves, and changes from hour to hour or day to day. Over many years, certain conditions become familiar weather in an area. The average weather in a specific region, as well as its variations and extremes over many years, is referred to as climate (http://www.nationalgeographic.org/encyclopedia/weather/). There are six main components, or parts of weather. They are temperature, atmospheric pressure, wind, humidity, precipitation, and cloud. These components together describe the weather at any given time. These shifting components, along with the knowledge of atmospheric processes, help meteorologists or scientists who study weather to forecast what the weather will be in the near future.

2.1.1 Rainfall

Rainfall is the most important source of moisture (water) (Ayodele, 2011). Rain is droplets of water that fall from clouds. Heat from the sun turns moisture (water) from oceans, lakes and rivers into water vapour (gas), which evaporate into the air. This vapour rises, cools, and changes into tiny water droplets which form clouds (Ibrahim, 2018). The most common rainfall measurement is the total rainfall depth during a given period, expressed in millimetres (mm). Rain gauge serves as device for collecting and measuring the amount of rain which fall over an area in predefined period of time. Rainfall is possibly the most important factor in defining climate and also the major source of energy that drives the circulation of the atmosphere. (Mitch, 1998)

2.1.2 Temperature

Temperature is the degree of sensible heat or cold. The distribution of temperature generally reflects the distribution of rainfall. Temperature has strong effect on rainfall as increase in temperature increases the evaporation and thus the moisture in the atmosphere increases which when reaches a sufficient level considering other parameters will cause rainfall. Temperature can be measured with an instrument called thermometer and it’s refer to how hot or cold the atmosphere is at given time (Ibrahim, 2018). Temperature can be measured in two ways: in Celsius (C) and Fahrenheit (F). The United States uses the Fahrenheit system while in other parts of the world, Celsius is used. Almost all scientists measure temperature using the Celsius scale. Temperature is a relative measurement. An afternoon at 70 degrees Fahrenheit, for example, would seem cool after several days of 95 degrees Fahrenheit, but it would seem warm after temperatures around 32 degrees Fahrenheit. The coldest weather usually happens near the poles, while the warmest weather usually happens near the Equator.

2.1.3 Atmospheric pressure

This is the weight of the atmosphere overhead. Changes in atmospheric pressure signal shifts in the weather. A high-pressure system usually bring cool temperatures and clear skies. A low-pressure system can bring warmer weather, storms, and rain. Atmospheric pressure changes with altitude. The atmospheric pressure is much lower at high altitudes. For example; the air pressure on top of Mount Kilimanjaro, which is 5,895 meters tall is 40 % of the air pressure at sea level (Pepin, 2017). The weather is much colder. The weather at the base of Mount Kilimanjaro is tropical, but the top of the mountain has ice and snow. Atmospheric pressures are measured in millibars or inches of mercury with an instrument called barometer. Average atmospheric pressure at sea level is about one atmosphere (about 1,013 millibars, or 29.9 inches). An average low pressure system, or cyclone, measures about 995 millibars (29.4 inches). A typical high-pressure system, or anticyclone, usually reaches 1,030 millibars (30.4 inches) (Ibrahim, 2018)

2.1.4 Wind

This is the movement of air parcels. Wind forms because of differences in temperature and atmospheric pressure between nearby regions. Winds tend to blow from areas of high pressure, where it's colder, to areas of low pressure, where it's warmer (Ibrahim, 2018). In the upper atmosphere, strong, fast winds called jet streams occur at altitudes of 8 to 15 kilometres (5 to 9 miles) above the Earth. They usually blow from about 129 to 225 kilometres per hour (80 to 140 miles per hour), but they can reach more than 443 kilometres per hour (275 miles per hour). These upper atmosphere winds help push weather systems around the globe. Wind speed can be measured with an instrument called anemometer while wind vanes measure wind direction (http://www.nationalgeographic.org/encyclopedia/weather/).

2.1.5 Humidity

Relative humidity is the amount of water vapour in the air. If the humidity is high, there is more moisture in the air and so more probabilities of clouds forming and rainfall. Water vapour is a gas in the atmosphere that helps make clouds, rain, or snow. Humidity is usually expressed as relative humidity, or the percentage of the maximum amount of water air can hold at a given temperature. Cool air holds less water than warm air. At a relative humidity of 100 %, air is said to be saturated, meaning the air cannot hold any more water vapour. Excess water vapour will fall as precipitation. Clouds and precipitation occur when air cools below its saturation point. This usually happens when warm, humid air cools as it rises. The most humid places on Earth are islands near the Equator. For example, Singapore is humid year round. The warm air is continually saturated with water from the Indian Ocean. Hygrometer is used to measure the amount of water vapour in air, in soil, or in confined spaces (Ibrahim, 2018).

2.1.6 Clouds

Clouds come in a variety of forms and shapes. Not all of them produce precipitation. Wispy cirrus clouds, for example, usually signal mild weather. Other kinds of clouds can bring rain or snow. A blanket-like cover of nimbo-stratus clouds produces steady, extended precipitation. Enormous cumulo-nimbus clouds, or thunderheads, release heavy downpours. Cumulo-nimbus clouds can produce thunderstorms and tornadoes as well. Clouds can affect the amount of sunlight reaching the Earth's surface. Cloudy days are cooler than clear ones because clouds prevent more of the sun's radiation from reaching the Earth's surface. The opposite is true at night then, clouds act as a blanket, keeping the Earth warm. Cloud patterns indicate the present of weather systems, which produce most of the weather we are familiar with; for examples, rain, heat waves, cold snaps, humidity and cloudiness (Ibrahim, 2018).

2.2 Rainfall Variation in Nigeria

Rainfall varies across space and time scale. In Nigeria there is variation across the various regions of the country and time. Most literatures on rainfall point to the fact that rainfall in Nigeria decreases as one move from the south (coastal region) towards the north (arid region) (Ngamdu, 2015). Rainfall in the Northern part of Nigeria has a unimodal distribution starting mostly in June-July and reaches its peak in August and end rapidly in September to October (Ayoade, 2004 and Adejuwon, 1990).

Oyewole *et al*. (2014), studied and analysed rainfall variation in Nigeria with thirty-one-year data from (1979-2010). Statistical package was used for the analysis, the mean; standard deviation and coefficient of variation were obtained. The results showed that Minna had highest rainfall in 2001 and lowest in 1987 in August, June, and July respectively. The result also showed that rainfall increases generally by 0.6546 per year in Nigeria.

Previous studies have analysed rainfall trends over entire or part of Nigeria. For example, (Adefolalu, 1986) examined trends in rainfall pattern using 70-year period (1911–1980) rainfall data from twenty-eight meteorological stations. Bello, (1998) extended the work and compared the seasonality of rainfall distribution in Nigeria for two climatic periods, 1930–1961 and 1962–1993. Ati *et al*. (2009), described the significant increase in rainfall over nine stations in northern Nigeria between 1953 and 2002. The results showed a general decline of dry season's contribution to annual rainfall, that is, dry period is getting drier. In the recent past, (Oguntunde, 2011) analysed rainfall trends over Nigeria using 1901–2002 rainfall data from global gridded climatology of Climate Research Unit Time Series (CRUTS). They observed that annual rainfall has been reduced significantly over 20 % of the landscape and the amount of annual rainfall reduced by 50–350 mm in 64 % portion of Nigeria. Present study differs from the previous ones in the following perspective; firstly, ground observation data used by Adefolalu, (1986) and Bello, (1998) was extended to year 1999. Secondly, the statistical approach employed is different and more recent.

The recent studies from different researchers aim to detect change (transition) point in rainfall pattern over Nigeria on a climatic zone basis using non-parametric statistical analysis. This was performed on the average zonal rainfall to show transitions in wet to dry and dry to wet among three (3) different non-overlapping climatic periods of 30 years each, that is, 1910–1939, 1940–1969 and 1970–1999. This enabled the observation of increase or decrease in rainfall received in each of the climatic zones of Nigeria to be established. There were upward and downward shifts during the study period over all the zones. At the Guinean zone, there was partial recovery in 1942 and lasted for about 30 years before another change point in 1970. At Savanna, there was relative dryness since early 1920 but transits in 1970. At the Sahel, there was downward and upward shift in 1950 and 1969 respectively. The change points were clearly significant over the Sahelian zone at 99 % confidence level, which agrees with (Oluleye, 2009) and (Ogungbenro and Morakinyo, 2014).

2.3 Related Works on Rainfall Prediction

In 2018, an over view of Seasonal Rainfall Prediction (SRP) was published by Nigerian Meteorology Agency (NIMET), produce the SRP to reduce the negative impact of prediction and to provide a vital tool for informed decision making, information and planning in rainfall and environmental management. The timely preparation and release of the SRP is to enhance preparedness against predicted climate risk and hazards. SRP is therefore an invaluable tool which is capable of ameliorating the unpleasant consequences of extreme weather and climate events which reduces climate related risk. The SRP results show that: (1) the country is expected to have normal rainfall in most places. (2) Rainfall amount above 3100 mm are likely expected in Eket and Calabar while Niger and Benue are expected to have 800 mm to 1600 mm, areas around Nguru, Yelwa and Ibi to have about 5-7% above normal rainfall (NIMET, 2018).

Using the Gumbel's extreme value theory, (Ologunorisa and Tersoo, 2006) estimated the return periods of extreme daily rainfall at Makurdi, Nigeria between 1979 and 2004. The result shows that the period between 1996 and 2001witnessed the highest frequencies of extreme rainfall events; that major floods were associated with higher return periods. In most cases, floods were associated with either abnormally high daily rainfall events or annual rainfall events (Umar, 2012). Ayoade, (1976) also examined the magnitude, frequency and distribution of intense rainfall in Nigeria using maximum daily rainfalls of forty-six Nigerian stations with varying duration ranging from 35 to 59 years. They computed the recurrence interval (return period) of the highest daily rainfall ever recorded at each of the forty-six stations using the Gumbel's Extreme Value theory. The results showed that the return period of the extreme daily rainfall decreases from the south towards the northern part of the country.

Abdulkadir *et al.* (2012) developed a real time Artificial Neural Network based on rainfall forecasting model for Ilorin, Nigeria, using observed rainfall records. This ANN model is designed to run a real time task in which the input to the model is a consecutive data of the rainfall. The neural network was trained with sixty years (1952–2011) total monthly historical rainfall data. The trained network yielded 76 % and 87 % of good forecast for the training and testing data set respectively. The correlation coefficient of 0.88 was obtained which showed that the network is fit to be used for the subsequent quantitative prediction of rainfall in Ilorin.

Olaiya and Adeyemo (2013), investigated the use of data mining techniques in predicting maximum temperature, rainfall, evaporation and wind speed. Decision tree algorithm and artificial neural networks were used for prediction. The meteorological data was collected between 2000 and 2009 from Ibadan, Nigeria. A data model for the meteorological data was developed and used to train the classifier algorithms. The performance of each algorithm was compared with the standard performance metrics and the algorithm with the best result was used to generate classification rules for the mean weather variables. A predictive neural network model was also developed for weather prediction and the results were compared with the actual weather data for the predicted period. The results show that given enough training data, data mining technique can be efficiently used for weather prediction and climate change studies.

Theethagiri *et al.* (2020) studied the different machine learning classification algorithms to predict the COVID-19 recovered and deceased cases. The k-fold cross-validation resampling technique was used to validate the prediction model. The prediction algorithm was evaluated with performance metrics such as prediction accuracy, precision, recall, mean square error, confusion matrix. The KNN algorithm predicts 92 % (true positive rate) of the deceased cases correctly, with 0.077 % of misclassification. Further, the KNN algorithm produces the lowest error rate as 0.19 on the prediction of accurate COVID-19 cases than the other algorithm. Also, it produces the receiver operator characteristic curve with an output value of 82 %.

Zhang and Zhou (2016), performed a comparative analysis of Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Adaptive Neuro Fussy Inference

System (ANFIS) on rainfall prediction. The authors compared the prediction models in four terms: (i) by using different lags as modelling inputs; (ii) by using training data of heavy rainfall events only; (iii) performance of forecasting for 1 hour to 6 hours and; (iv) performance analysis in peak values and all values. According to results obtained by the authors, ANN performed better when trained with dataset of heavy rainfall. For 1 to 4 hour ahead forecasting, the previous 2-hour input data was suggested for all three modelling techniques (ANN, SVM and ANFIS). ANFIS reflected better ability in avoiding information noise by using different lags of inputs. Finally, during peak values, SVM proved to be more robust under extreme typhoon events.

Julie and Kannan (2010), considered an approach to handle learning disability database to predict frequent signs and symptoms of the learning disability in school age children. Two classification techniques, decision tree and clustering was used for 125 real data sets with most of the attributes takes binary values the data, where each partition represents a cluster and it classifies the data into groups. Each group contains at least one object and each object must belong to exactly one group. The results obtained the accuracy of the classifier was 77.6 % and incorrectly classified 22.4 %. Sensitivity of 84% and area under curve (AUC) of 71%.

Furthermore, Tyagi and Kumar (2017) predicted monthly rainfall by using Back Propagation, Radial Basis Function and Neural Network. For prediction, the dataset was collected from Coonoor region in Nilgiri district (Tamil Nadu). Performance was evaluated in terms of Mean Square Error. According to the results, higher accuracy was reported in Radial Basis Function Neural Network with smaller Mean Square Error. Moreover, the researchers also used these techniques for future rainfall prediction. However, devastating events such as flooding mostly caused by heavy rainfall are natural phenomenon that cannot be stopped from occurring but its effects can be minimised if effectively tackled and appropriate measures are taken to slow down its effects and frequency (Bukka *et al*. 2017).

2.4 Flooding

Flooding in Nigeria and other parts of the world has been reported to affect and render people homeless than any other natural disaster (Bukka *et al*., 2017). According to Ladds (2017), flooding can simply be described as "water where it is not wanted". It can also be conceptualised as a situation that results when a part of the earth surface that is usually dry is inundated and covered with water due to high amount of rainfall or the overflowing of a water body. Among different natural disasters, flooding causes severe damages to properties. There are two basic types of floods: flash floods and the more widespread river floods. Flash floods generally causes greater loss of life and river floods generally causes greater loss of property.

2.4.1 Flash floods

This type of flood occurs when runoff from excessive rainfall causes a rapid rise in the water height (stage) of a stream or normally-dry channel. Flash floods are more common in areas with a dry climate and rocky terrain because lack of soil or vegetation allows torrential rains to flow over land rather than infiltrate into the ground.

2.4.2 River floods

This type of flood is generally more common for larger rivers in areas with a wetter climate, when excessive runoff from longer-lasting rainstorms and sometimes from melting snow causes a slower water-level rise over a larger area. Floods also can be caused by ice jams on a river or high tides, but most floods can be linked to a storm of some kind. Etuonovbe (2011) believed that 20 % of the world populace faces the risk of flooding caused by heavy rainfall and that it is dangerous to communities. Flooding tends to inundate the farmlands and makes them waterlogged. Douglas *et al.* (2005)

expressed that agricultural sector being the main source of livelihood in agrarian communities is the worst affected by flooding in most flood ravaged communities in the world. Flooding also amplifies the transport problems due to flooded roads and dented infrastructure especially the 2012 flooding in Niger Delta, Nigeria, where Niger Deltans were displaced, forcing many inhabitants out of their homes and halting economic progress (Ajayi *et al.*, 2012). Many lives and properties worth millions of naira got lost because of flooding.

According to Niger State Emergency Management Agency (NSEMA, 2018) at least fourteen persons died due to flooding in different parts of Minna and over sixty houses were affected by flood due to heavy rainfall in Kontagora, Tafa and Suleja Local Government Areas of the State. On 9th September 2018, according to The Guardian Newspaper, over thirty villages were affected by flood in Mokwa Local Government Area of the State, Houses, farmlands and live stocks were destroyed in this disaster (https://www.vanguardngr.com/2018/07/over-60-houses-affected-by-flood-in-niger-nsema/).

2.5 Concept of Artificial Neural Network

Artificial neural networks (ANNs) are computational model that consists of several processing elements that receive inputs and deliver outputs based on their predefined activation functions. They are biologically inspired computational networks (Park and Lek, 2016). Artificial neural networks take inspiration from the learning process occurring in human brains. They consist of an artificial network of functions called parameters, which allows the computer to learn by analysing new data. These parameters, sometimes also referred to as neurons, are function that yields an output, after receiving one or multiple inputs. Those outputs are then passed to the next layer of neurons, which use them as inputs of their own function, and produce further outputs.

Those outputs are then passed on to the next layer of neurons, and so it continues until every layer of neurons have been considered, and the terminal neurons have received their input. Those terminal neurons then output the final result for the model. The ability of the neural networks to learn is stimulated in the same way in which the human brain works. Human brain is a highly complex, nonlinear, and parallel information-processing system. Neural Networks are simplified models of biological neuron system (Nayak, 2013). An artificial neural network (ANN) models can be describe as the structural and the functional aspects of biological neural networks mathematically. The neural networks are a complex adaptive system, which changes its own internal structure with the help of the information, which is flowing through it. Generally, these changes can be done by altering the weights, which is assigned to each neuron. Figure 2.1 shows a visual representation of such a network. The initial input is *x*, which is then passed to the first layer of neurons where three functions consider the input that they receive, and generate an output. The output is then passed to the second layer. There, further output is calculated based on the output from the first layer. That secondary output is then combined to yield a final output of the model.

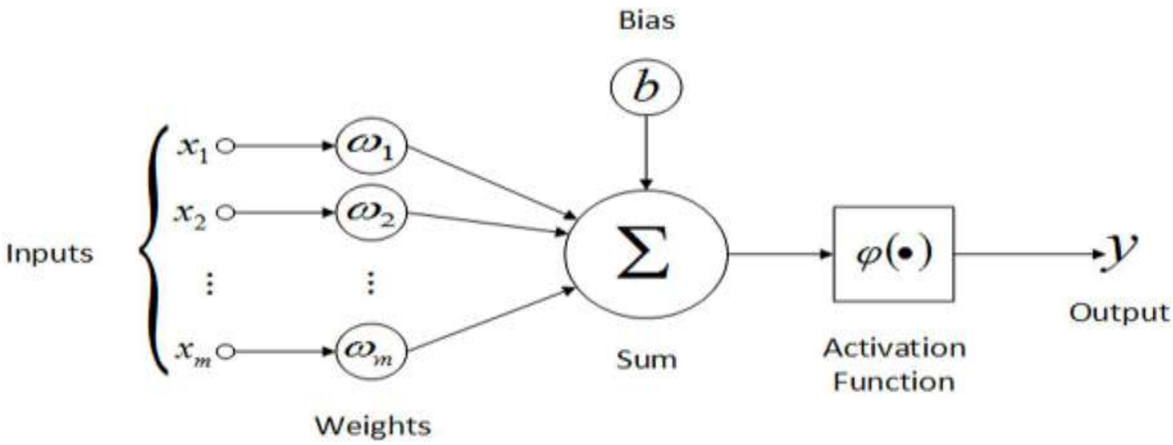

Figure 2.1: Structure of ANN

2.5.1 Structure of artificial neural network

ANNs are easy to construct and deal very well with large amounts of noisy data. They are mainly suited to solving nonlinear problems. They work well for problems where domain experts may be unapproachable or where there are no known rules. ANNs are also adaptive in nature. This makes them mostly suitable in many fields where the environment is potentially volatile and dynamic (Ahire, 2018). They are also very tolerant of noisy and incomplete data sets. The basic structure of an ANN consists of artificial neurons (similar to biological neurons in the human brain) that are grouped into layers. The most common ANN structure consists of an input layer, one or more hidden layers and an output layer. A modified simple model of an artificial neuron is shown in Figure 2.2

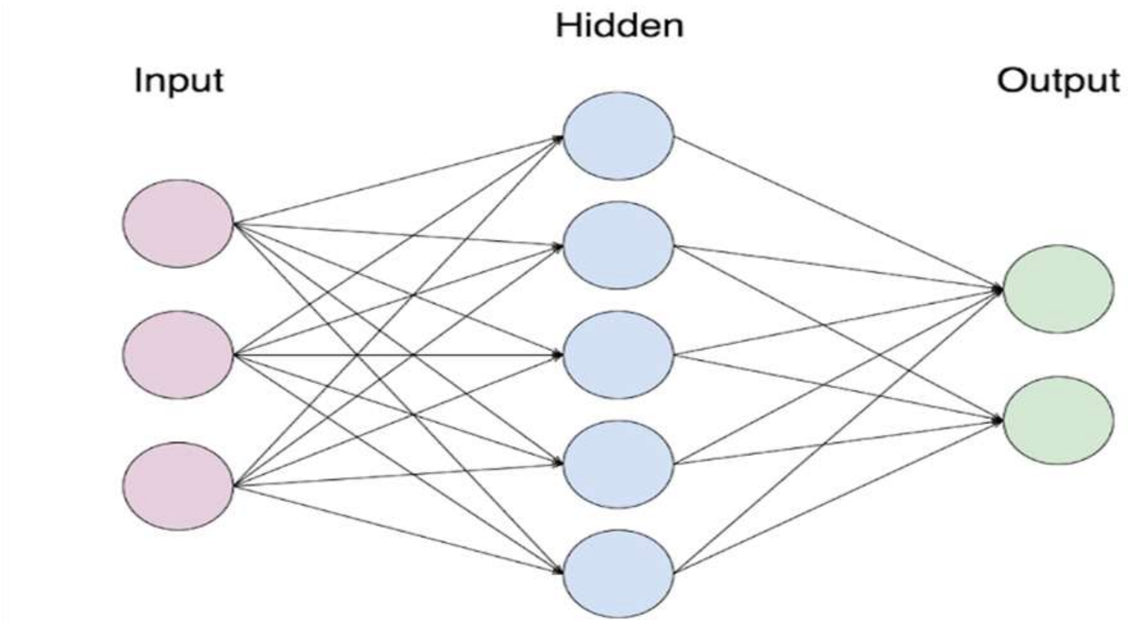

Figure 2.2 Visual representation of a simple neural net

The terms associated with Figure 2.2 are explained as follows:

Input layer: This is the first layer in the neural network. It takes input data and passes them on to the next layer. It doesn't apply any actions on the input values and has no weights and biases value associated.

Hidden layer: Hidden layers have neurons which apply different transformation to the input values. This layer is responsible for the calculations that are performed within the network. A hidden layer can have more than one neuron.

Output layer: this layer is the last layer in neural network and receives the input data from the last hidden layer. With this layer, the desired output data and desired range are available at this layer. In some cases, a neural network can have more than one output layers.

In the human brain, neurons communicate by sending signals to each other through complex connections. ANNs are based on the same principle in an attempt to simulate the learning process of the human brain by using complex algorithms. Every connection has a weight attached which may have either a positive or a negative value associated with it. Positive weights activate the neuron while negative weights constrain it. In Figure 2.1, a network structure with more than one inputs being connected to more than one hidden neuron on each connection was shown. The neuron sums all the signals it receives, with each data being multiplied by its associated weights on the connection.

$$\boldsymbol{z} = \sum \boldsymbol{x_1} * \boldsymbol{w_1} + \boldsymbol{x_2} * \boldsymbol{w_2} + \ldots + \boldsymbol{x_n} * \boldsymbol{w_n} + \boldsymbol{b} * \boldsymbol{1} \tag{2.1}$$

This output (y) is then passed through a transfer (activation) function, *g(y) that* is normally non-linear to give the final output, j. The most commonly used function is the sigmoid (logistic function) because of its easily differentiable properties, which is very convenient when the back-propagation algorithm is applied.

$$\hat{y} = sigmoid(z) \tag{2.2}$$

$$sigmoid(z) = \frac{1}{1+e^{-z}} \tag{2.3}$$

where y = model output, z = model input, e = exponential function.

Generally, the working of a human brain by making the right connections is the idea behind ANNs. That was restricted to use of silicon and wires as living neurons and dendrites. Here, neurons (part of human brain) that was composed of 86 billion nerve cells, also connected to other thousands of cells by axons. Although, there are various inputs from sensory organs that are accepted by dendrites. As a result, it creates electric impulses that is used to travel through the ANN. Thus, to handle the different issues, neuron sends a message to another neuron. As a result, it can be said that ANNs composed of multiple nodes that imitates biological neurons of the human brain. Also, they interact with each other. Although, nodes are used to take input data further perform simple operations on the data. As a result, these processes are passed to other neurons. Also, output at each node is called its activation or node value as each link is associated with weight. Also, they are capable of learning (https://data-flair.training/blogs/artificial-neural-network/)

2.6 Classification Prediction

The classifiers are used to find the class to which an unknown data belongs based on the information available from a set of data whose class is already known. Classification models predict categorical class labels; and prediction models predict continuous valued functions. Classification refers to a predictive modelling problem where a class label is predicted for a given sample of input data (Brownlee, 2020). From a modelling perspective, classification requires a training dataset with many samples of input and output from which to learn. A model will use the training dataset and calculate how to best map instances of input data to specific class labels. As such, the training dataset must be sufficiently representative of the problem and have many sample of each class label. The class label is often string values, for example “rain”, “not rain”, and must be mapped to numeric values before being provided to an algorithm for modelling. This is often referred to as label training, where a unique integer is assigned to each class label, for example, “rain” = 1, ‘no rain” = 0. Classification predictive modelling algorithms are evaluated based on their results. Classification accuracy is a popular metric used to evaluate the performance of a model based on the predicted class label.

Classification aims at capturing characteristics of training data in order to distinguish between the dataset and potential outliers to appear. One classifier is fit on a training dataset that only has example from the normal class. Once prepared, the model is use to classify new examples as either normal or not-normal (i.e. outliers or anomalies).

There are four major types of classification modelling. This includes (a) binary, (b) multi, (c) multi-label and (d) imbalanced classification (Brownlee, 2020).

2.6.1 Binary classification

Binary classification refers to those classification tasks that have two class labels. Examples include: rainfall prediction (rain or not), email spam detection (spam or not), conversion prediction (buy or not). Typically, binary classification tasks involve one class that is the normal state and another class; the abnormal state. For example, “not spam” is the normal state and “spam” is the abnormal state. Another example is “cancer not detected” is the normal state of a task that involves a medical test and “cancer detected” is the abnormal state. The class for the normal state is assigned the class label

0 and the class with the abnormal state is assigned the class label 1. It is common to model a binary classification with a model that predicts a Bernoulli probability distribution for each example. The Bernoulli distribution is a discrete probability distribution that covers a case where an event will have a binary outcome as either 0 or 1. For classification, this means that the model predicts a probability of an example belonging to class 0 or 1 (i.e. normal state or the abnormal state).

Activation functions are very important feature of Artificial Neural Network. They define the output of a node given an input or set of inputs. That is, activation functions calculate a "weight sum" of its input, adds a bias and then decides the output. Binary step function is one of the activation functions in Artificial Neural Network popular known as "threshold function". All the input, x_i is multiplied with their weight w_i assigned to each link and summed together along with bias, b. The gradient of binary step function is zero, which is not suitable in back propagation for weight update. It can handle binary class problem only. Although with some tweaks, it can be used for multi-class problems. Figure 2.3 shows binary step function.

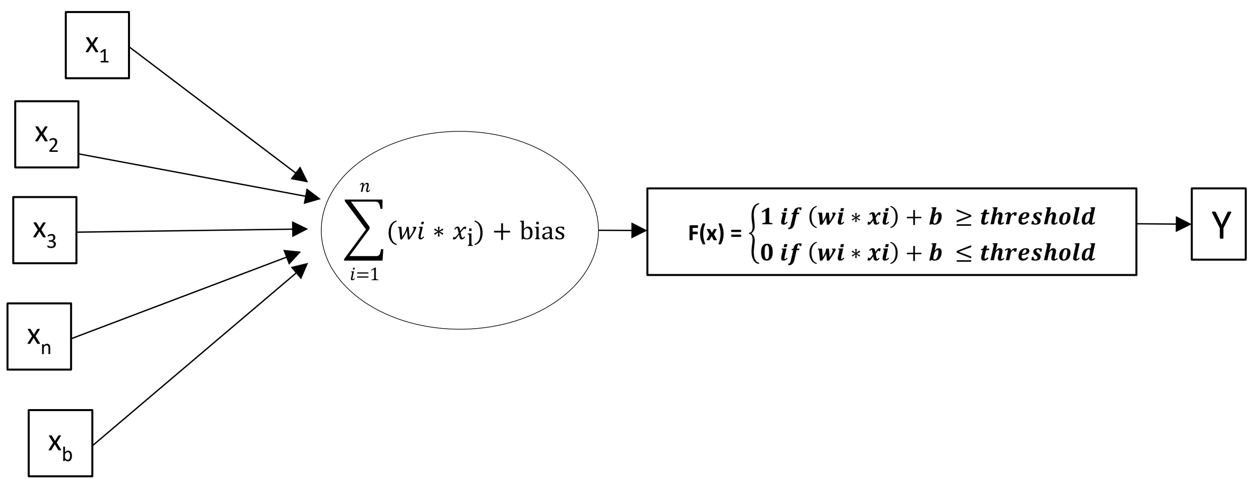

Figure 2.3: Binary Step Function

2.6.2 Multi-classification

Multi-class classification refers to those classification tasks that have more than two class labels. Examples include: Face classification, Plant species classification, Optical character recognition. Unlike binary classification, multi-class classification does not have the notion of normal and abnormal outcomes. Instead, examples are classified as belonging to one among a range of known classes. The number of class labels may be very large on some dataset. For example, a model may predict a photo as belonging to one among thousands or tens of thousands of faces in a face recognition system.

2.6.3 Multi-label classification

Multi-label classification refers to those classification tasks that have two or more class labels, where one or more class labels may be predicted for each dataset. Consider the example of photo classification, where a given photo may have multiple objects in the scene and a model may predict the presence of multiple known objects in the photo, such as "bicycle," "apple," "person," etc. This is unlike binary classification and multi-class classification, where a single class label is predicted for each example.

2.6.4 Imbalanced classification

Imbalanced classification refers to classification tasks where the number of examples in each class is unequally distributed. Typically, imbalanced classification tasks are binary classification tasks where the majority of examples in the training dataset belong to the normal class and a minority of examples belong to the abnormal class. Examples include: Fraud detection, outlier detection, Medical diagnostic tests.

The step by step process for calculating a confusion Matrix in data mining is; first test the dataset with its expected outcome values, then predict all the rows in the test dataset. Calculate the expected predictions and outcomes, count the total of correct predictions of each class and also count the total of incorrect predictions of each class. Classification is one of the most popular machine learning technique to predict the class of new samples, using a model inferred from training data. In general, classification is defined as a learning method that maps or classifies data instances into the corresponding class labels that are predefined in the given dataset. According to Han *et al*. (2011), data classification is a two-step process; first one is the learning step where a classification model is constructed from a given dataset; the data from which a classification function or model is learned is known as the training set, and second one is a classification step where the model is used to test or predict the class labels for a separate unseen given data; the data set that is used to test the classifying ability of the learned model or function is known as the testing set (Sarker *et al.*, 2019).

CHAPTER THREE

3.0 MATERIALS AND METHODOLOGY

In this chapter, the method applied in carrying out this work is discussed. The section consists of flowchart of research design, data collection, data preparation approach and Network features description.

3.1 Materials

3.1.1 Data

For this research, four (4) years meteorological data of Rainfall (mm), Relative Humidity, Minimum and Maximum Temperature (°C) for Minna, was collected from the Geography Department, Federal University of Technology Minna for the period 2015 to 2018.

3.2.1 Data analysis

The data were collected on daily and annual total basis. 870 (that is, 70 %) datasets with three attributes were documented and analysed for training phases, while 186 (that is, 15 %) dataset documented for testing and 186 (that is, 15 %) dataset documented for validation phases. This implies that 1242 dataset were used for this work.

3.2.2 Data preparation

The prime objective of data preparation is to minimise the missing values, outliers, fields that are obsolete or redundant, data in a form not suitable for model, normalisation (scaling) for values not consistent with policy or common logic that gets into the network so as to reduce the error in the network. According to Pyle, (1999) data preparation alone accounts for 60% of all the time and effort expanded in the entire data mining process.

3.2.3 Missing data

Missing data is one of the problems that continue to plague data analysis methods. Even as analysis methods gain sophistication, missing values will be encountered especially in databases with a large number of fields. In this work, the missing values were replaced with the fields mean and for large gap, the fields were filtered.

3.2.4 Data scaling (Normalisation)

The dataset tends to have ranges that vary greatly from each other and such differences in the ranges will lead to a tendency for the variable with greater range to have undue influence on the results (Larose, 2005). Therefore, it is important to normalise numerical variables to standardise the scale for good results. There are two major techniques for normalisation: Min-Max normalisation and Z-score standardisation. In this work, Min-max normalisation was used. This normalisation technique works by seeing how much greater the field value is than the minimum value min (x) and scaling this difference by the range. That is,

$$\dot{\mathrm{X}} = \frac{x - \min(x)}{\operatorname{range}(x)} \tag{3.1}$$

where

x = refer to the original field value, $\dot{\mathbf{X}}$ is the normalised field value, min(x) is minimum field value and range(x) is the difference between maximum field value and minimum field value. This means, Min–Max normalisation values will range from 0 to 1, unless new data values are encountered that lie outside the original range.

3.2.5 Data splitting

It is ethical to split the dataset to avoid over fitting or under fitting. The splitting proportions are decided according to the size and type of the data available (Kumar, 2020). In this work, splitting ratio of 70:15:15 was used to split the dataset, that is, the dataset was split into three parts: 70% for train data, 15% for validation data and 15%

testing data. The training set has to be largest because large data sample is used to fit the network. The Artificial Neural Network observes and learns from this data and optimises its parameters.

3.3 Rainfall prediction

Feed Forward Neural Network and Binary classification was used to predict the unknown class (new rain days) which was compared with the true class label (that is, already known rainy days). The target data (that is, rainfall) was label 1 and 0 (rain and no rain). The prediction process involve fitting the Artificial Neural network classifier on a training dataset and making prediction on a test dataset. The predictions are in the form of normalised probabilities. In order to use binary classification, a threshold must be defined (classification threshold). The default threshold is always 0.5, but thresholds are problem dependent. In this work, a value above threshold mapped to class 1, indicates “rain day”; value below threshold mapped to class 0, indicates “no rain day”. Different threshold values were then tried and the class labels were evaluated using confusion matrix. The threshold that achieves the best evaluation is the adopted for the validating dataset when making prediction on new data in the future.

3.4 Rainfall Prediction Analysis

Confusion matrix was used to analyse the prediction outcomes. In binary classification prediction, there are four major prediction outcomes that could occur as shown in Table 3.1 and further simplified in Table 3.2.

Table 3.1 Confusion matrix

N	Predicted No rain	Predicted Rain
No rain	TN	FP
Rain	FN	TP

Table 3.2 Confusion matrix outcome: FN/TN/FP/TP

Actual rain day	Predicted rain day	Outcomes
0	1	False Positive
0	0	True Negative
1	0	False Negative
1	1	True Positive

where: N = total number of the test dataset sample.

True Positive (TP): this means that the actual target is positive (1) and the outcome of the prediction is positive (1). In relation to this work, True Positive implies that there was rainfall in the actual day and there will be rainfall in predicted day.

True Negative (TN): this means the actual target is negative (0) and the outcome of the prediction is negative (0). True Negative implies that there was no rainfall in the actual day and there will be no rainfall in the predicted day.

False Positive (FP): this means the actual target is negative (0) and the outcome of the prediction is positive (1). False Positive implies that there was no rainfall in the actual day and there will be rainfall in the predicted day.

False Negative (FN): this means the actual target is positive (1) and the outcome of the prediction is negative (0). False Negative implies that there was rainfall in the actual day and there will be no rainfall in the predicted day.

These four outcomes were used to evaluate the classifier accuracy, error, sensitivity, specificity.

3.5 Prediction Evaluation

The four outcomes of the prediction analysis were used to evaluate the classifier. The evaluation standards are discussed below:

3.5.1 Accuracy

Accuracy is defined as the percentage of correct predictions of the rain days and no rain days. It can be calculated easily by dividing the number of correct predicted rain days and no rain days by the number of total predicted days.

$$accuracy = \frac{TP+TN}{N} \tag{3.2}$$

3.5.2 Sensitivity

Sensitivity is defined as the percentage of predicted rain days. This mean when it’s actually rain, how often the classifier correctly predicts rain days.

$$sensitivity = \frac{TP}{TP+FN} \tag{3.3}$$

3.5.3 Specificity

Specificity is defined as the percentage of predicted no rainfall days. This implies that when it’s actually no rain day, how often the classifier correctly predicts no rainfall day.

$$specificity = \frac{TN}{TN+FP} \tag{3.4}$$

3.5.4 Misclassification

Misclassification is defined as the percentage of wrong prediction of the rain days and no rain days. It can be calculated by dividing the number of wrong predictions of rain days and no rain days by the number of total predicted days. It is also known as error rate

$$error\ rate = \frac{FN+FP}{N} \tag{3.5}$$

3.6 Prediction Performance Evaluation

It is necessary to evaluate the prediction classifier. Receiver Operating Characteristics curve (ROC curve) is a plot that show the diagnostic capacity of a binary classification prediction as its threshold is varied. It is often used to explain the connection between sensitivity and specificity for every possible threshold. For a given choice of threshold, the accuracy can be computed; it is the proportion of sensitivity and false positive rate. The higher the area under the curve (AUC) the better the classifier, that is, Area under the curve close to 1 indicates a good performance. It helps to find the best classifier and threshold that represents data and how well the chosen threshold will work in the future to make good and accurate predictions. It also aims to estimate the generalisation accuracy of a classifier.

3.7 Flowchart of the Method

The flowchart implemented for this work is shown in Figure 3.1

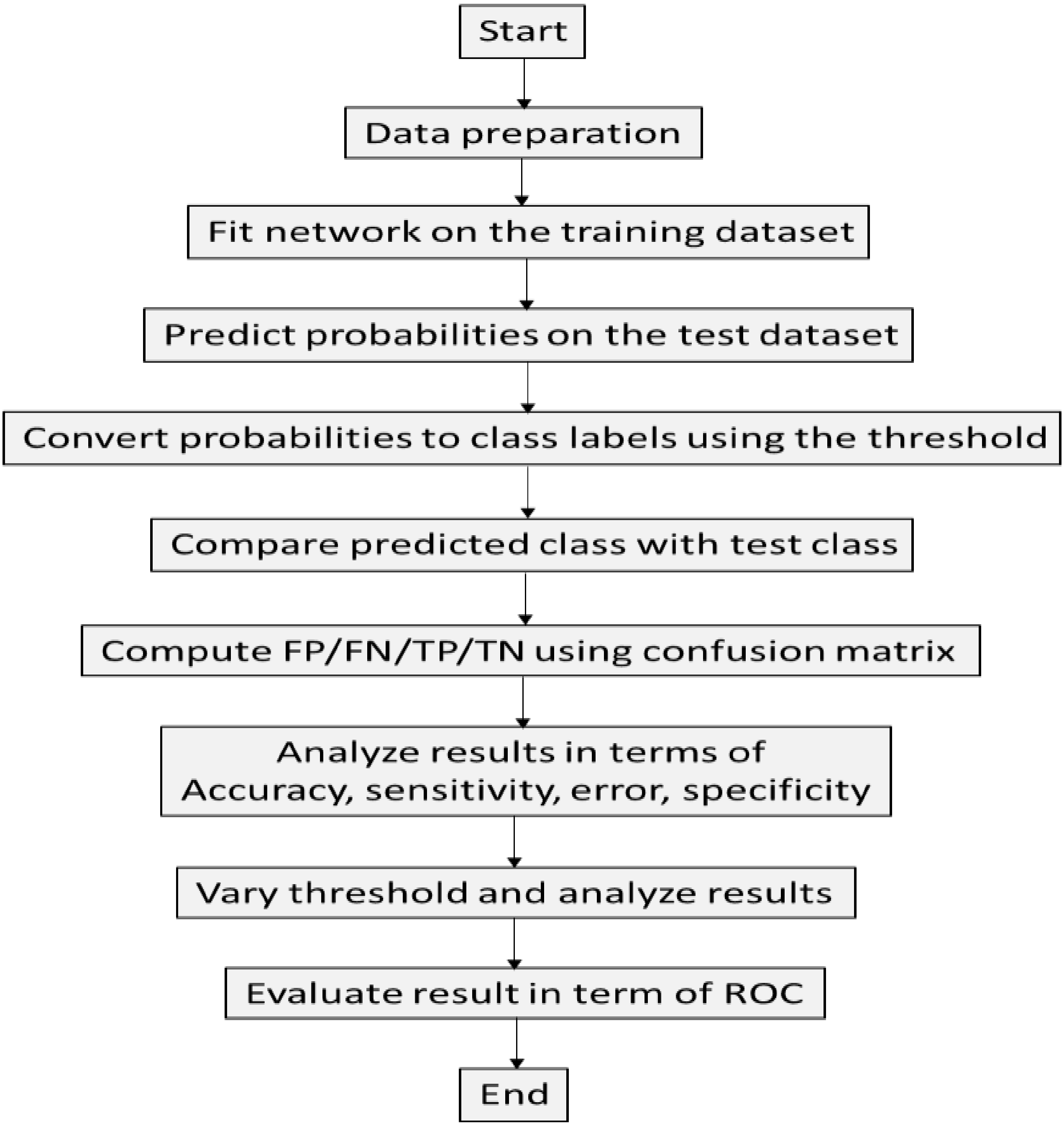

Figure 3.1: Flowchart of the research methodology

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

A study of the rainfall variation is necessary for an informed prediction. This work is mainly about rainfall prediction of the study area, therefore the variation of rainfall over Minna metropolis was analysed. The monthly and yearly rainfall variation was computed by summing all the amount of rainfall obtained for a raining month and year respectively.

4.1. Monthly Rainfall Variation

The monthly rainfall variation for the year 2015 is shown in Figure 4.1. The rainfall spanned from April to September with May having the highest rainfall of 258.9 mm and September having the lowest rainfall of 10.2 mm. The dry season spanned from January to March and from October to December with no rainfall.

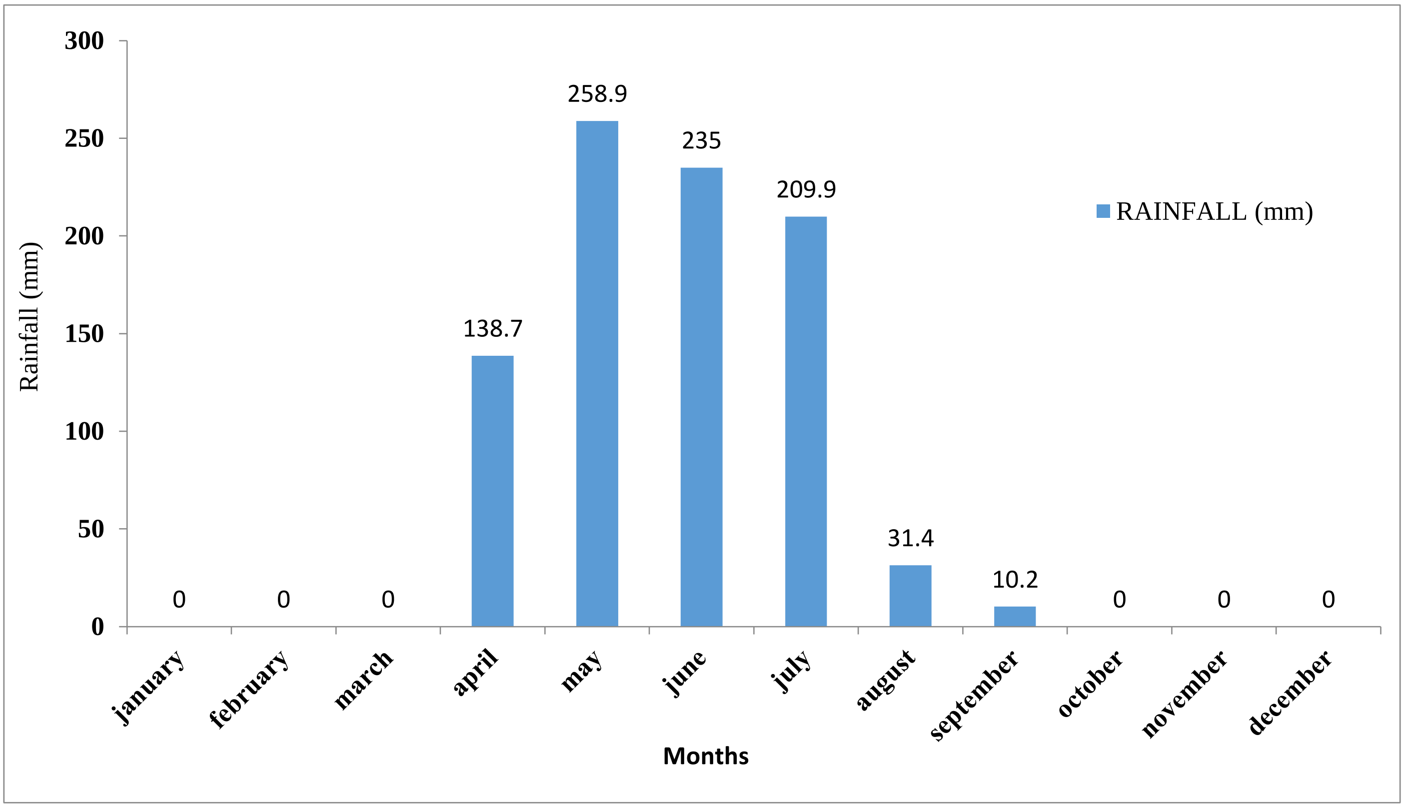

Figure 4.1: Monthly distribution of rainfall for year 2015

The monthly rainfall variation for the year 2016 is presented in Figure 4.2. The dry months was from January to February and December with no rainfall. Rainfall began from March to November with September having the highest rainfall of 320.8 mm and March having the lowest rainfall of 15 mm.

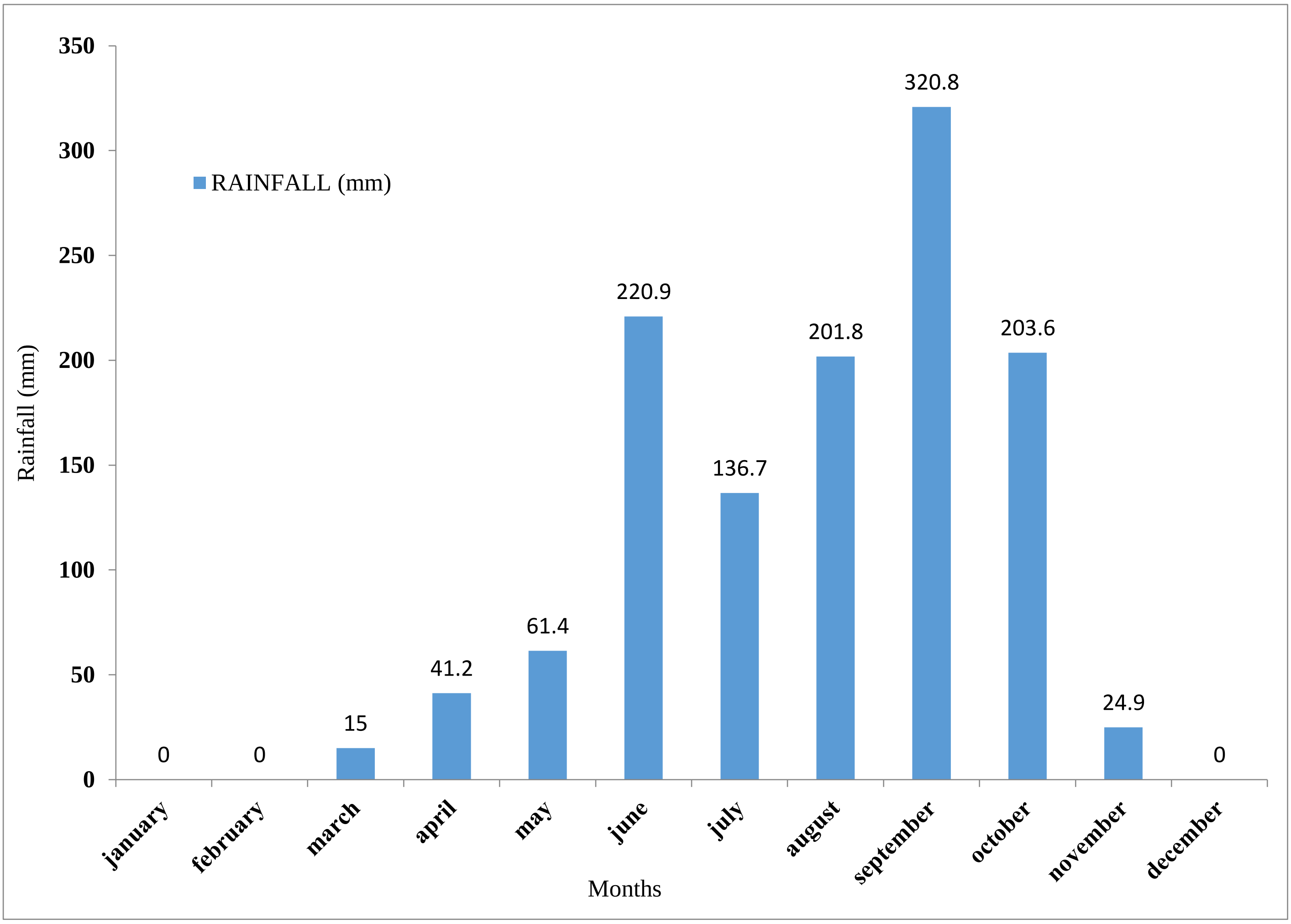

Figure 4.2: Monthly distribution of rainfall for year 2016

Figure 4.3 shows the monthly variation for the year 2017. The dry months are from January to March and from November to December with no rainfall while the rainfall began from April to October with July having the highest rainfall of 253.3 mm and April with low rainfall of 8.5 mm.

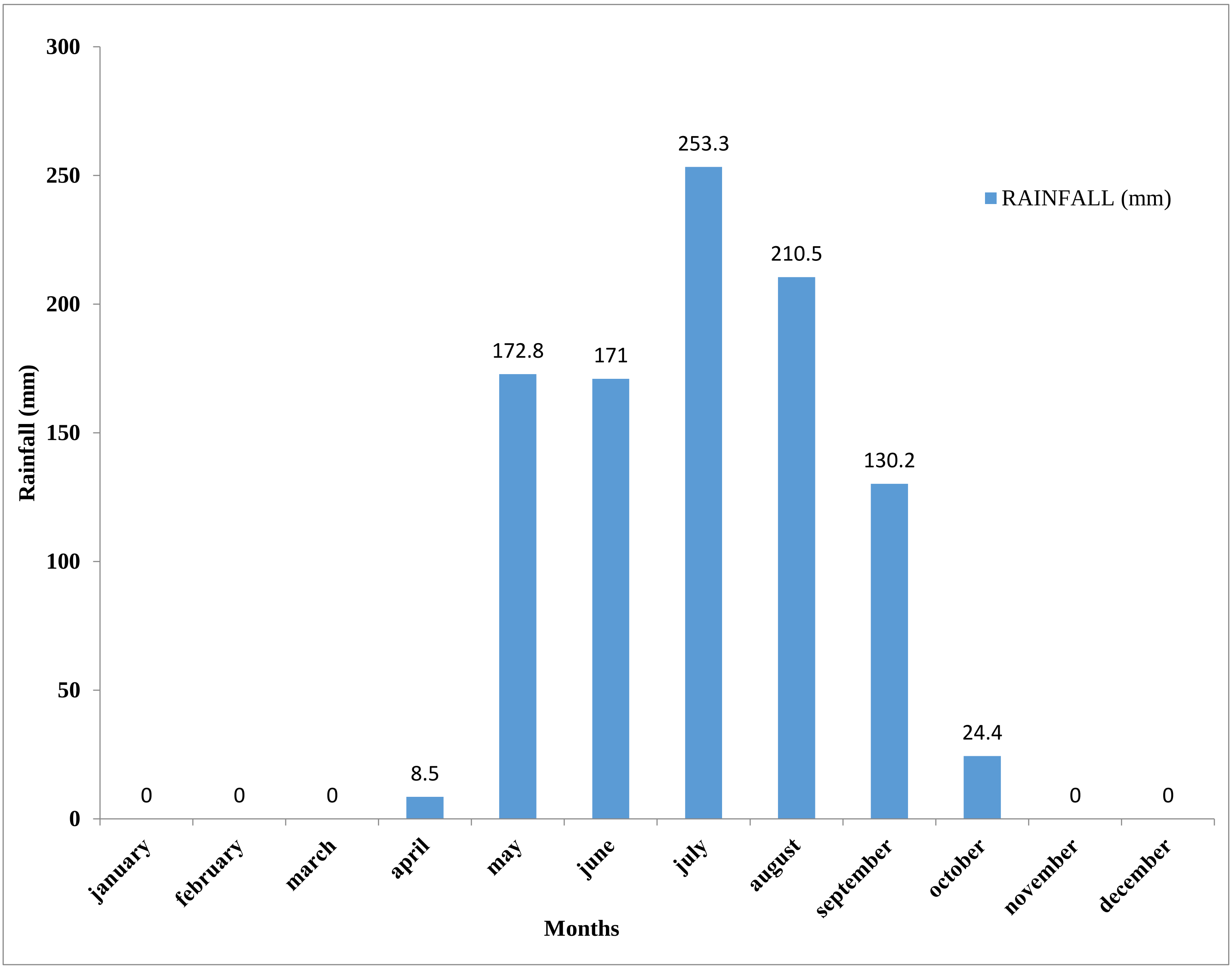

Figure 4.3: Monthly distribution of rainfall for year 2017

Figure 4.4 shows the monthly variation for the year 2018. The rainfall began from April to October with August having the highest rainfall of 294.2 mm and April with lowest rainfall of 25.1 mm. Also, the dry months began from January to March and November to December with no rainfall.

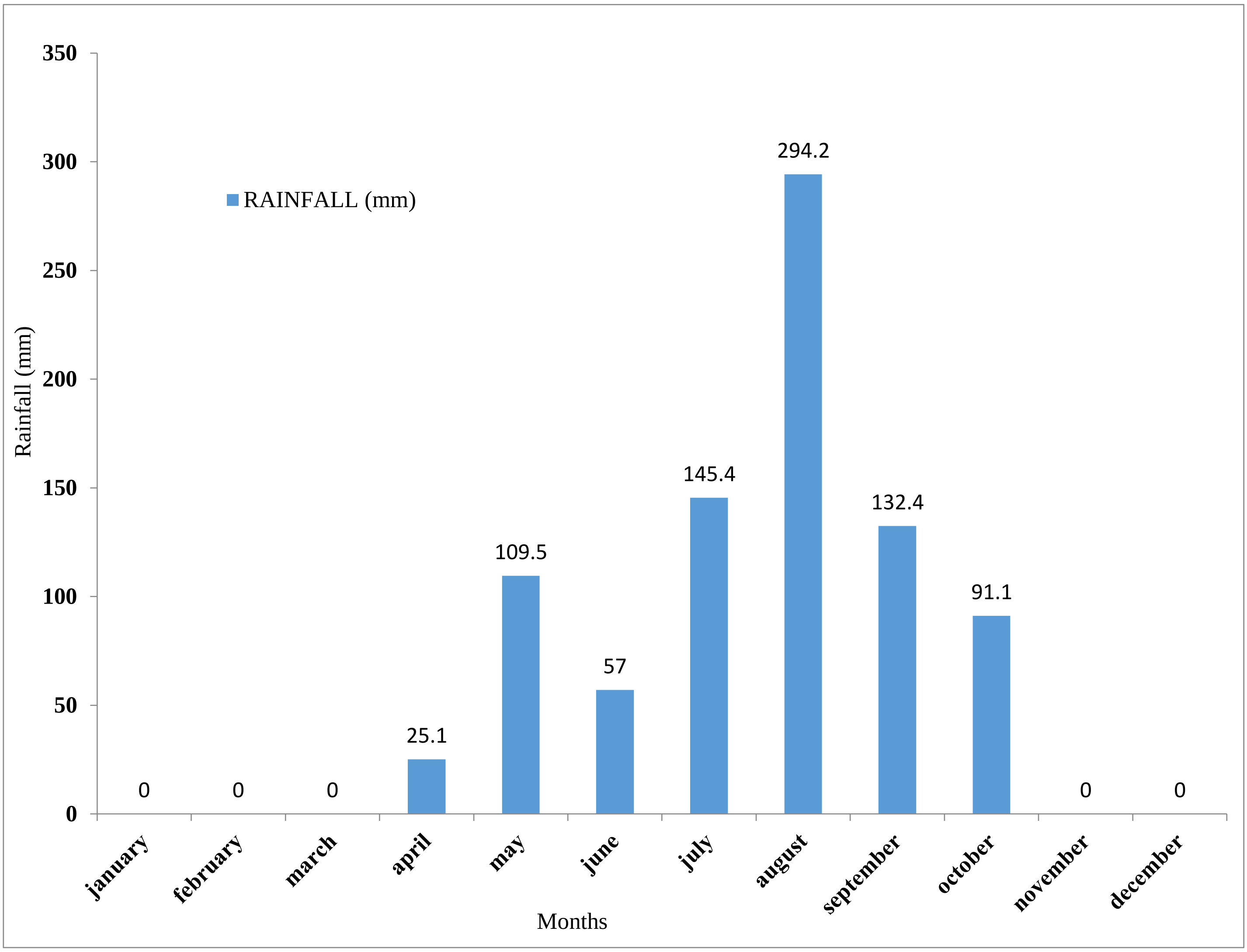

Figure 4.4: Monthly distribution of rainfall for year 2018

Figure 4.5 shows the comparison of the monthly rainfall distribution from years 2015 to 2018. September 2016 has the highest monthly rainfall with 320.8 mm, followed by August 2018 with 294.2 mm monthly rainfall while May 2017 has the lowest monthly rainfall with 8.5 mm followed by November 2015 with 10.2 mm monthly rainfall.

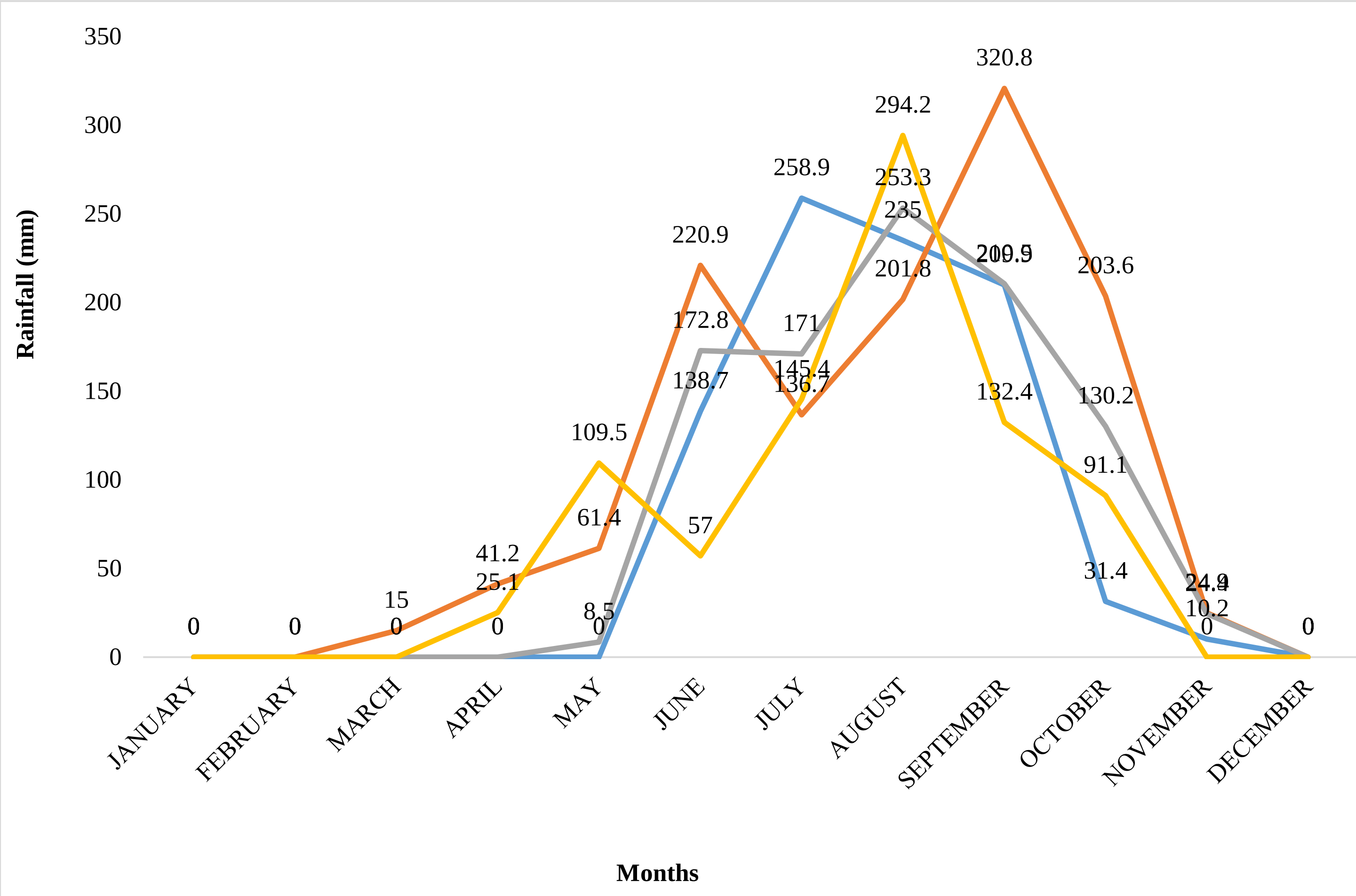

Figure 4.5: Comparison of the monthly rainfall distribution for the study periods

4.1.2 Yearly rainfall variation

The yearly or annual rainfall variation for the years under consideration was computed by summing the rainfall of the constituent months. Figure 4.5 shows the result of the rainfall variation for the years 2015 to 2018. Year 2016 has the highest rainfall of 1226.3 mm followed by 2017 with 970.7 mm, 2015 with rainfall of 884.1 mm and 2018 with a rainfall of 854.7 mm which is the least for the years under consideration.

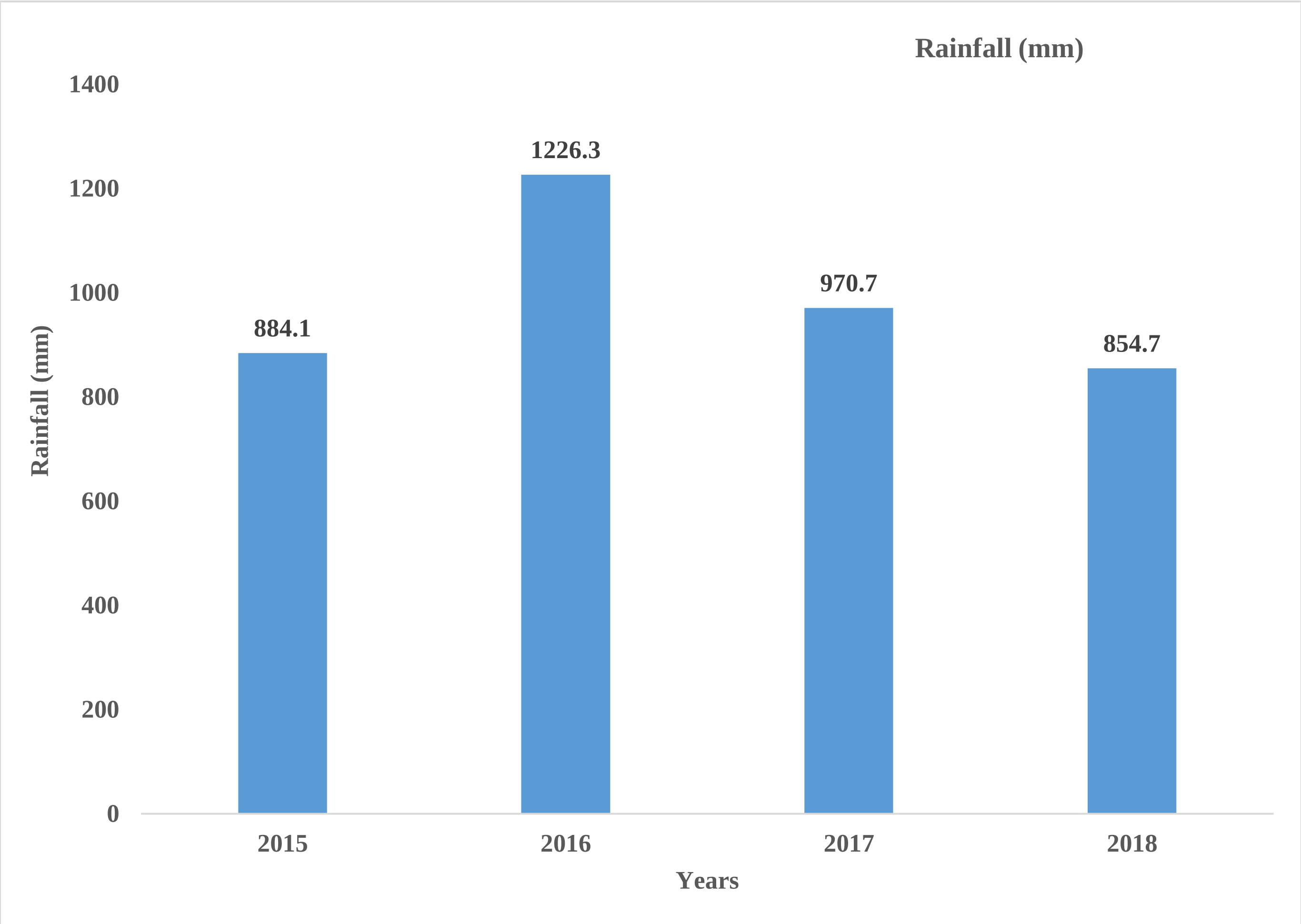

Figure: 4.6 Yearly rainfall variations (2015-2918)

The rainfall variation for the four-year data shows that rainfall varies each year. These variations can be as results of change in temperature, a common factor affecting rainfall and other weather related variables of the rainfall. The rainfall variation examined in this study agreed with Ayoade (2004) and Adejuwon (1990) analysis, that Minna tends to have rainfall (wet season) from July to September and dry season from December to March. Furthermore, according to Oyewole *et al.* (2014), Minna experiences highest rainfall in the months of August followed by June and July in the year 2001 and 1987.

4.2 Rainfall Prediction

4.2.1 Data pre-processing stages

Data pre-processing is a technique that involves cleaning and transforming raw data into an understandable format. Figures 4.6 and 4.7 shows the sample of raw data obtained and normalised data for this research work. Figure 4.6 captures from January 1st, 2015 to July 17th, 2018. Raw data can be defined as unanalysed data, that is; data not yet subject to analysis. The dataset consists of four year atmospheric parameters (maximum temperature, minimum temperature, relative humidity and rainfall), the whole dataset total four columns and 1313 rows. The raw dataset contains some missing field values which can affect the training if subjected to the Artificial Neural Network without proper pre-processing. To minimise error, the raw data collected was pre-processed and normalised. The missing field values were replaced with the mean values and those with wide range of empty fields were filtered away from the dataset. This process reduced the size of original dataset to 1242 rows from 1313 rows in the raw data. The pre-processed data was normalised. After implementing equation (3.1) on the processed data, the data became normalised. The Min-max normalisation values ranges from 0 or

1 as shown in Table 4.2. Table 4.1 shows screenshot of raw data before pre-processing while Table 4.2 shows screenshot preprocessed and normalised results respectively.

Table 4.1: Sample of raw data

Raw Data					Raw Data				
Year/Date	Minimum Temperature(°C)	Maximum Temperature (°C)	RH(%)	Rainfall (mm)	Year/Date	Minimum Temperature(°C)	Maximum Temperature (°C)	RH(%)	Rainfall (mm)
01/01/2015	23	35	30	0	16/06/2018	22	29	93.5	0
02/01/2015	21	35	30	0	17/06/2018	23	30	93.5	0
03/01/2015	21	35	30	0	18/06/2018	23	30	96	0.5
04/01/2015	21	35	44	0	19/06/2018	22	30	96	0
05/01/2015	21	35	32	0	20/06/2018	23	30.5	91	0
06/01/2015	19	35	40	0	21/06/2018	22	31	87	0
07/01/2015	20	32	38	0	22/06/2018	22	33	91.5	0
08/01/2015	20	31	29	0	23/06/2018	22	31	91.5	0
09/01/2015	21	31	33	0	24/06/2018	21	32	91.5	0
10/01/2015	20	31	36	0	25/06/2018	21	32	96	14.5
11/01/2015	18	29	34	0	26/06/2018	21	31	91	17.5
12/01/2015	20	30.1	35	0	27/06/2018	23	32	96	0
13/01/2015	19	31	29	0	28/06/2018	22	33	83	0
14/01/2015	24	34	32	0	29/06/2018	23	32	96	7.5
15/01/2015	21	34	30	0	30/06/2018	22	32	96	0
16/01/2015	21	35	36	0	01/07/2018	23	32	96	0
17/01/2015	21	35	23	0	02/07/2018	22	32	96	1
18/01/2015	22	36	34	0	03/07/2018	23	32	96	5.2
19/01/2015	21	35	48	0	04/07/2018	22	31	96	30.2
20/01/2015	21.5	35	73	0	05/07/2018	22	31	96	0
21/01/2015	23	37	63	0	06/07/2018	22	31	96	0
22/01/2015	23.5	37	56	0	07/07/2018	22	31	96	0
23/01/2015	21.5	39	29	0	08/07/2018	22	31	97	0
24/01/2015	22	37	52	0	09/07/2018	23	31	96	22.5
25/01/2015	23	35	78	0	10/07/2018	23	32	96	0
26/01/2015	23	38	82	0	11/07/2018	23	32	96	0
27/01/2015	23	38	48	0	12/07/2018	23	30	96	14.5
28/01/2015	23	37	51	0	13/07/2018	23	30	91	5.2
29/01/2015	23	37	91	0	14/07/2018	23	32	93	0
30/01/2015	24	37	91	0	15/07/2018	23	32	94	0
31/01/2015	24	37	91	0	16/07/2018	25	33	96	0
01/02/2015	23	38	49	0	17/07/2018	25	32	80	15

Table 4.2: Sample of normalised data

Processed and normalised Data				Processed and normalised Data			
Minimum Temperature(°C)	Maximum Temperature (°C)	RH(%)	Rainfall (mm)	Minimum Temperature(°C)	Maximum Temperature (°C)	RH(%)	Rainfall (mm)
0.318181818	0.565217391	0.1428571	0	0.454545455	0.086956522	0.725274725	0
0.318181818	0.565217391	0.1428571	0	0.454545455	0.608695652	0.802197802	0
0.318181818	0.565217391	0.2967033	0	0.409090909	0.608695652	0.813186813	1
0.318181818	0.565217391	0.1648352	0	0.409090909	0.608695652	0.813186813	0
0.227272727	0.565217391	0.2527473	0	0.409090909	0.608695652	0.824175824	0
0.272727273	0.434782609	0.2307692	0	0.409090909	0.565217391	0.813186813	1
0.272727273	0.391304348	0.1318681	0	0.409090909	0.47826087	0.824175824	0
0.318181818	0.391304348	0.1758242	0	0.727272727	0.434782609	0.78021978	1
0.272727273	0.391304348	0.2087912	0	0.590909091	0.391304348	0.868131868	1
0.181818182	0.304347826	0.1868132	0	0.454545455	0.347826087	0.824175824	1
0.272727273	0.352173913	0.1978022	0	0.727272727	0.52173913	0.868131868	0
0.227272727	0.391304348	0.1318681	0	0.454545455	0.347826087	0.769230769	1
0.454545455	0.52173913	0.1648352	0	0.454545455	0.130434783	0.824175824	0
0.318181818	0.52173913	0.1428571	0	0.409090909	0.304347826	0.659340659	1
0.318181818	0.565217391	0.2087912	0	0.727272727	0.434782609	0.868131868	0
0.318181818	0.565217391	0.0659341	0	0.590909091	0.434782609	0.824175824	0
0.363636364	0.608695652	0.1868132	0	0.590909091	0.391304348	0.769230769	1
0.318181818	0.565217391	0.3406593	0	0.545454545	0.434782609	0.824175824	0
0.340909091	0.565217391	0.6153846	0	0.409090909	0.391304348	0.813186813	1
0.409090909	0.652173913	0.5054945	0	0.5	0.391304348	0.824175824	0
0.431818182	0.652173913	0.4285714	0	0.545454545	0.434782609	0.824175824	0
0.340909091	0.739130435	0.1318681	0	0.5	0.434782609	0.813186813	1
0.363636364	0.652173913	0.3846154	0	0.318181818	0.173913043	0.857142857	1
0.409090909	0.565217391	0.6703297	0	0.363636364	0.391304348	0.813186813	0
0.409090909	0.695652174	0.7142857	0	0.431818182	0.434782609	0.813186813	1
0.409090909	0.695652174	0.3406593	0	0.409090909	0.304347826	0.769230769	0
0.409090909	0.652173913	0.3736264	0	0.409090909	0.391304348	0.868131868	1
0.409090909	0.652173913	0.8131868	0	0.409090909	0.434782609	0.813186813	0
0.454545455	0.652173913	0.8131868	0	0.454545455	0.391304348	0.774725275	0

4.2.2 Training, testing and validating stages

In this work, 70:15:15 ratio was used to split the dataset. 1242 data samples were subjected to the Artificial Neural Network for training, testing and validation. 70 % for training, 15 % for testing and 15 % for validating corresponding to 870, 186 and 186 samples respectively. The training data was used to fit the classifier and testing data was used to test the classifier. The validation data was used to predict new set (output) based on the training.

4.2.3 Prediction stage

Binary classification prediction method of artificial neural network was used to predict the rainfall days and no rainfall days. The test data was fed into the already trained network. Threshold of 0.5 was used to classify which days will have rainfall and days with no rainfall. The predicted days are shown in the 3rd column of Table 4.3. The table also compared the actual days (2nd column) with the predicted days (3rd column).

Recalling Table 3.1, the four outcomes which are True positive, True Negative, False Positive and False Negative were all worked out and presented. Table 4.3 is the compilation of the rainfall prediction and confusion matrix outcomes.

Table 4.3: Sample of rainfall prediction results

DAYS	TRUE LABEL	PREDICTED LABEL	CM
1	0	1	FP
2	0	0	TN
3	0	0	TN
4	0	0	TN
5	1	0	FN
6	0	0	TN
7	0	0	TN
8	0	0	TN
9	1	0	FN
10	0	0	TN
11	0	0	TN
12	1	0	FN
13	1	1	TP
14	0	0	TN
15	1	0	FN
16	1	1	TP
17	0	0	TN
18	1	1	TP
19	1	0	FN
20	1	0	FN
21	1	1	TP
22	0	1	FP
23	1	1	TP
24	0	0	TN
25	0	1	FP
26	0	1	FP
27	0	1	FP
28	1	1	TP
29	0	1	FP
30	1	1	TP
31	1	1	TP
32	0	1	FP
33	0	1	FP
34	0	1	FP
35	0	1	FP
36	1	1	TP
37	1	1	TP
38	0	1	FP
39	1	0	FN
40	1	1	TP
41	0	1	FP
42	1	1	TP
43	0	1	FP
44	0	0	TN

DAYS	TRUE LABEL	PREDICTED LABEL	CM
45	0	0	TN
46	0	0	TN
47	0	0	TN
48	0	1	FP
49	0	0	TN
50	0	0	TN
51	0	0	TN
52	0	0	TN
53	0	0	TN
54	0	0	TN
55	0	0	TN
56	0	0	TN
57	0	0	TN
58	0	0	TN
59	0	0	TN
60	0	0	TN
61	0	0	TN
62	0	0	TN
63	0	0	TN
64	0	0	TN
65	0	0	TN
66	0	0	TN
67	0	0	TN
68	0	0	TN
69	0	0	TN
70	0	0	TN
71	0	0	TN
72	0	0	TN
73	0	0	TN
74	0	0	TN
75	0	0	TN
76	0	0	TN
77	0	0	TN
78	0	0	TN
79	0	0	TN
80	0	0	TN
81	0	0	TN
82	0	0	TN
83	0	0	TN
84	0	0	TN
85	0	0	TN
86	0	0	TN
87	0	0	TN
88	0	0	TN

89	0	1	FP	135	0	0	TN
90	0	0	TN	136	0	0	TN
91	1	1	TP	137	0	0	TN
92	0	0	TN	138	0	0	TN
93	0	0	TN	139	0	0	TN
94	0	1	FP	140	0	0	TN
95	1	1	TP	141	0	0	TN
96	0	0	TN	142	0	0	TN
97	0	1	FP	143	0	0	TN
98	0	0	TN	144	0	0	TN
99	0	1	FP	145	0	0	TN
100	0	1	FP	146	0	0	TN
101	0	1	FP	147	0	0	TN
102	0	1	FP	148	0	0	TN
103	0	1	FP	149	0	0	TN
104	0	1	FP	150	0	0	TN
105	0	1	FP	151	1	0	FN
106	0	1	FP	152	0	0	TN
107	0	1	FP	153	0	0	TN
108	0	1	FP	154	0	0	TN
109	0	1	FP	155	0	0	TN
110	0	1	FP	156	0	0	TN
111	0	1	FP	157	0	0	TN
112	0	1	FP	158	0	0	TN
113	0	1	FP	159	0	0	TN
114	0	1	FP	160	0	0	TN
115	1	1	TP	161	0	0	TN
116	0	0	TN	162	0	0	TN
117	0	0	TN	163	0	0	TN
118	1	1	TP	164	0	0	TN
119	0	0	TN	165	1	1	TP
120	0	1	FP	166	0	0	TN
121	0	1	FP	167	0	0	TN
122	0	1	FP	168	0	0	TN
123	0	1	FP	169	0	0	TN
124	1	1	TP	170	0	0	TN
125	0	1	FP	171	0	0	TN
126	0	1	FP	172	0	0	TN
127	0	0	TN	173	1	0	FN
128	0	0	TN	174	0	0	TN
129	0	0	TN	175	0	0	TN
130	0	0	TN	176	1	0	FN
131	0	0	TN	177	0	0	TN
132	1	0	FN	178	0	0	TN
133	0	0	TN	179	1	0	FN
134	0	1	FP	180	0	0	TN

181	0	0	TN	184	1	0	FN
182	0	1	FP	185	0	0	TN
183	0	0	TN	186	1	0	FN

The outcome of table 4.3 is the first round results obtained. It is expected that several rounds of results should be obtained by re-running the test data. The respective outcome will be analysed and best results extracted. The procedure as stated was carried out for nine more times. A summary of the confusion matrix obtained by summing the number of P, N, TP, FP, TN and FN for each round of the ten rainfall prediction results was compiled and presented in Table 4.4

Table 4.4: Summary of rainfall for ten iterations prediction

S/N	P	N	TP	FP	TN	FN
1	59	127	18	41	113	14
2	29	157	10	19	135	22
3	25	161	13	12	142	19
4	66	120	20	46	108	12
5	32	154	17	15	139	15
6	35	151	14	21	133	18
7	42	144	13	29	125	19
8	53	133	16	37	117	16
9	38	148	12	26	128	20
10	73	113	16	57	97	16

The four possible outcomes of the classifier as computed using confusion matrix can be interpreted as: P (Positive) means predicted rain days, N (negative) means predicted no rain days, TP (True Positive) means both the actual day and predicted day have rainfall, TN (True Negative) means actual day has no rainfall and predicted day has no rainfall as well, FP (False Positive) means no rainfall in actual day but predicted day has rainfall, FN (False Negative) means actual day has rainfall but predicted day has no rainfall.

4.3 Prediction Analysis

The objective of the prediction analysis is to help identify which of the ten rainfall predictions made is most accurate. The analysis involves those of Accuracy, Error, Sensitivity and Specificity earlier presented in equation (3.1) to (3.4) respectively and are hereby recalled:

$$\boldsymbol{Error\ rate = \frac{FN + FP}{N}}$$

$$\boldsymbol{Accuracy = \frac{TP + TN}{N}}$$

$$\boldsymbol{Specificity = \frac{TN}{TN + FP}}$$

$$\boldsymbol{Sensitivity = \frac{TP}{TP + FN}}$$

The results of rainfall prediction analysis are presented in Table 4.5

Table 4.5 Rainfall prediction analysis

S/N	P	N	ACCURACY	ERROR	SENSITIVITY	SPECIFICITY
1	59	127	70%	1.3%	56%	86%
2	29	157	78%	2.2%	31%	93%
3	25	161	83%	2.6%	41%	92%
4	66	120	69%	1.3%	63%	84%
5	32	154	84%	2.0%	53%	89%
6	35	151	79%	1.9%	44%	90%
7	42	144	74%	1.7%	41%	91%
8	53	133	72%	1.4%	50%	88%
9	38	148	75%	1.8%	38%	91%
10	73	113	61%	1.3%	50%	86%

Sensitivity is the probability of a particular day to have rainfall while specificity is the probability of a particular day with no rainfall. The higher the sensitivity, the higher the probabilities of rainfall for those days, the 4th prediction results have the highest sensitivity of 63%. The accuracy tells how often the network classifies correctly and error is how often the network misclassifies. The 4^{th} prediction result has accuracy of 69% and error of 1.3% which ranked amongst the lowest. Sensitivity values is the best indicator since it gives the probability of having rainfall, a factor which this work is interested in predicting. 66 days were predicted with rainfall and 120 days with no rainfalls with sensitivity of 69 % and specificity of 84 %. The test target used for the prediction has 32 days of rainfalls and 154 days with no rainfalls. The predictions mean 34 days short in both cases.

4.4 Prediction Using Multiple Threshold

Having concluded that the 4^{th} iteration in Table 4.5, gives the best rainfall prediction with sensitivity of 69% and accuracy of 84%, it is also necessary to state that, a default threshold of 0.5 was used to classify which days will have rainfall and which days will not. In other to find out if there is a threshold that will perform better than 0.5, the threshold was therefore varied from 1.0 to 0.1. Table 4.6 shows the results of the random threshold selection.

Table 4.6: Multiple threshold

TV	P	N	TP	TN	FP	FN	ACCURACY (%)	SENSITIVITY (%)	SPECIFICITY (%)	FPR (%)
0.1	169	17	32	17	137	0	0.26	1.0000	0.3469	0.65306
0.2	110	76	30	74	80	2	0.56	0.9375	0.7115	0.28846
0.3	87	99	27	94	60	5	0.65	0.8738	0.7769	0.22314
0.4	60	126	20	114	40	12	0.73	0.7250	0.8507	0.14925
0.5	30	156	15	139	15	17	0.83	0.4688	0.9026	0.09740
0.6	11	175	2	145	9	30	0.79	0.0625	0.9864	0.01361
0.7	6	180	1	149	5	31	0.81	0.0313	0.9933	0.00667
0.8	4	182	2	152	2	30	0.83	0.0625	0.9870	0.01299
0.9	0	186	0	154	0	32	0.83	-	1.0000	-
1	0	186	0	154	0	32	0.83	-	1.0000	-

The performance of the various thresholds was evaluated using the yardsticks earlier presented. A graph of sensitivity against False Positive Rate (1-specificity) known as Receiver Characteristics Operating curve was plotted. The Receiver Characteristics Operating (ROC) curve is shown in Figure 4.7.

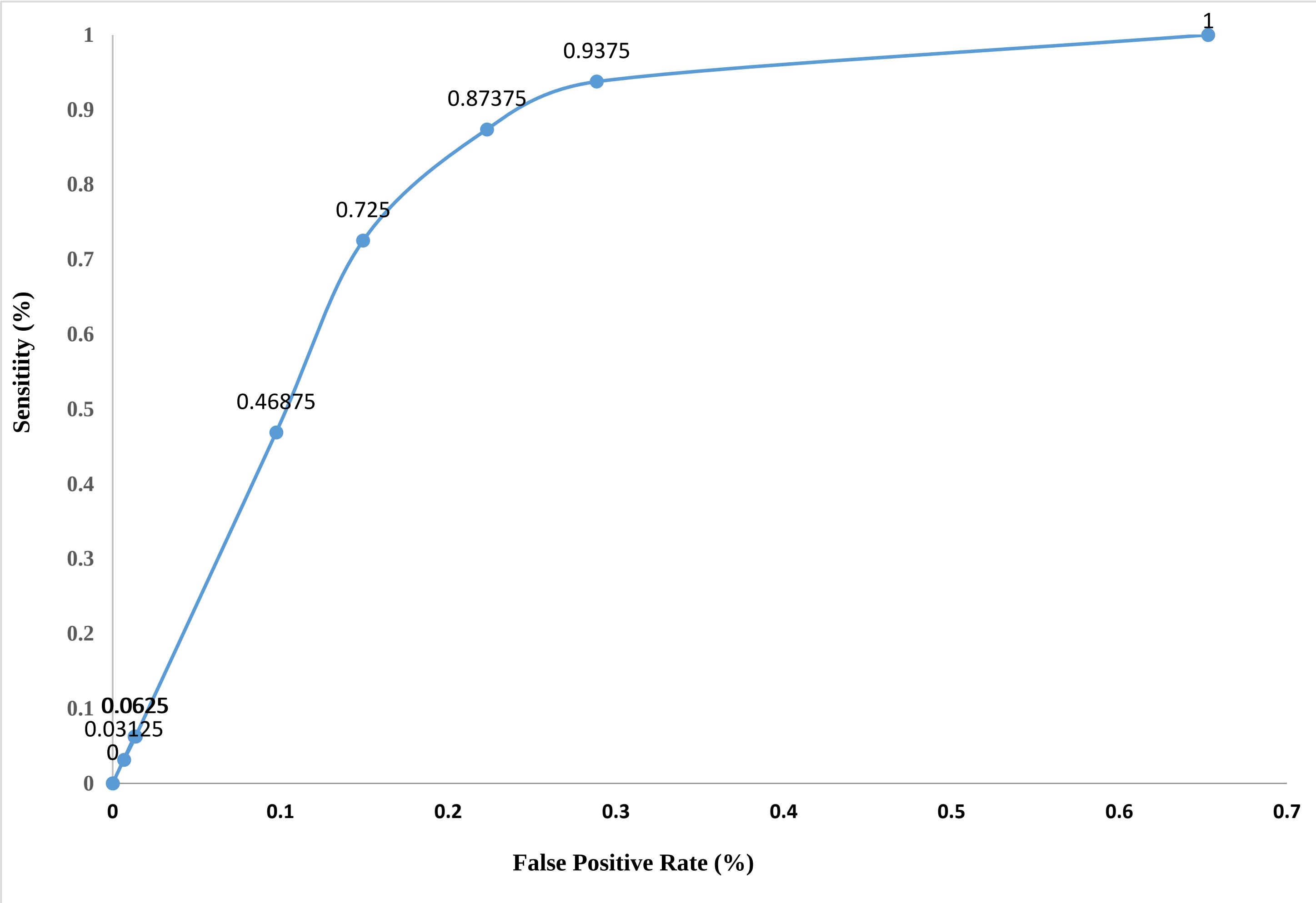

Figure 4.7: ROC curve

From Figure 4.7, the point on the ROC curve which is nearest to the left hand side (sensitivity) corner is 0.725 (72%). From Table 4.6, 72 % correspond to a threshold of 0.4. At this threshold, 60 rainfall days and 126 no rainfall days were predicted correctly out of 186 days in target (rainfall) data, with an accuracy of 73 %. Once again sensitivity is paramount, 72 % sensitivity rating of 0.4 threshold against that of 63 % of 0.5 threshold indicates that with a 0.4 threshold selection, the network predicts (classifies) rainfall better.

Validating target consist of 186 days, 46 days with rainfall (Positive) while 140 days with no rainfall (Negative). Threshold 0.4 was used to predict future rainfall days using validating target. Five iterations were made to predict future rainfall days and no rainfall days. Table 4.7 shows the predicted results.

Table 4.7: Validating prediction results

S/N	P	N	TP	TN	FP	FN	ACCURACY (%)	SENSITIVITY (%)	SPECIFICITY (%)	ERROR (%)
1	70	116	23	93	47	23	0.623656	0.5	0.801724	1.489361702
2	54	132	19	105	35	27	0.666667	0.413043478	0.846774	1.771428571
3	63	123	20	97	43	26	0.629032	0.434782609	0.82906	1.604651163
4	69	117	23	94	46	23	0.629032	0.5	0.803419	1.5
5	77	109	24	87	53	22	0.596774	0.52173913	0.783784	1.41509434

Out of five iterations, only the 5^{th}, 4^{th}, and 1^{st} have high sensitivity of 52 %, 50 % and 50 % respectively. For 5^{th} iteration (best prediction), 77 days were predicted as rain day while 109 days were predicted as no rain days with accuracy of 60 % and error of 1.4%. This means that extra 31 days were predicted to have rainfall with probability of 60%.

The Area under the ROC curve was computed which further shows the effectiveness of the classifier under three categories. These categories are excellent, good or not good using the criteria shown in Table 4.8. The area under the curve was calculated by applying equation (4.1) using an Excel formula. The area under the curve (AUC) having value close to 1 indicates a very good classifier.

$$AUC = ((A1-A2) * (B1) \qquad (4.1)$$

where,

A represent the false positive rate column and B represent the sensitivity column. Equation 4.1 was worked out and shown in Table 4.8.

Table 4.8: Area under the curve

S/N	FPR (%)	SENSITIVITY (%)	AUC (%)
1	0.653061	1	0.3646
2	0.288462	0.9375	0.061238
3	0.22314	0.87375	0.064559
4	0.149254	0.725	0.037592
5	0.097403	0.46875	0.03928
6	0.013605	0.0625	0.000434
7	0.006667	0.03125	-0.0002
8	0.012987	0.0625	0.000812
9	0	0	0
10	0	0	0
	AUC		0.568317

The value obtained for the AUC was compared with the conditions of the classifier performance rating shown in Table 4.9

Table 4.9: Classifier performance rating

AREA UNDER CURVE	CLASSIFIER
0.6 > 1.0	Excellent (very good)
0.50 > 0.59	Good
0 < 0.49	Not good (weak)

It can be observed that the AUC value of 0.56832 % correspond to the interval 0.5>0.59. Therefore, the rainfall classification (prediction) done in this work can be said to be in the “Good” categories

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMEDATIONS

5.1 CONCLUSION

This work predicts rainfall occurrence in Minna metropolis using Artificial Neural Network. However, the following conclusion were drawn: The monthly and yearly rainfall variations were examined, the findings from the rainfall variations shows that wet seasons began from May to October and dry seasons began from November to March respectively of the succeeding year. These variations were due to the change in related weather parameters especially temperature. For default threshold which is 0.5, 66 days were predicted to have rainfall with 69 % accuracy, 63 % sensitivity and 84 % specificity respectively. The threshold was varied and 73 % optimal accuracy was observed at 0.4 threshold. 0.4 threshold was used for validation, 77 days were predicted to have rainfall with 60 % accuracy, 52 % sensitivity, 78 % specificity and 1.4 % error respectively. This means that extra 31 days were predicted to have rainfall with probability of 60 %. It can be said that a threshold of 0.4 predicts rainfall better for Minna metropolis. This result is in the "good" category due to the fact that only four-year data was used.

5.2 RECOMMENDATIONS

The following recommendations and suggestions for further works are presented:

1. More and relevant atmospheric parameters such as wind, dew, average temperature among others should be incorporated into the dataset for a more encompassing rainfall prediction.
2. For higher accuracy of rainfall prediction in Minna using Artificial Neural Network, meteorological data for at least thirty years should be acquired. The

performances of Artificial Neural Network can be improved when input data is well prepared and when more data is employed for initial training.

3. With sufficient data incorporated, Artificial Neural Network can also be used to obtain a prediction model which is capable of predicting not only the occurrence of rainfall but the amount of rainfall.
4. Also with more data captured, 356 days can be used for testing and validation. These 356 days corresponds to the number of days in a year. Therefore, an annual rainfall prediction calendar can be published in advance.

REFERENCES

Abdulkadir, T. S., Salami, A. W & Kareem, A. G. (2012). Artificial neural network modeling of rainfall in Ilorin, Kwara State, Nigeria. *Journal of Research Information in Civil Engineering.*

Adefolalu, D. O. (1986). Rainfall Trends in Nigeria. Theoretical and Applied Climatology, 37(4), 205–219.

Adejuwon, J. (1990): Crop- climate relationship: the example of cocoa in Western Nigeria. *Nigerian Geography Journal, vol.5 pp 21-31.*

Aftab, S. & Ahmad, M. (2018). Rainfall prediction using data mining techniques: A systematic literature review (IJACSA) *International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5.*

Ahire, J. (2018). Artificial Neural Networks: The brain behind AI. Received from: Lulu.com

Ahmad M., Aftab S., (2017). Analyzing the performance of SVM for polarity detection with different datasets. *International Journal of Modern Education and Computer Science, 9(10), 29-36.*

Ajayi, O., Aboola, S. B. & Okesusim, F. B. (2012). Hydrology for Disaster Management. Special Publication of Nigeria Association of Hydrological Science. http://www.unaab.edu.ng

Akinsanola A. A & Ogunjobi K. O (2014). Analysis of rainfall and temperature variability over Nigeria. *Global Journal of Human-Social Science: B Geography, Geo-Sciences, Environmental Disaster Management. 2249-4604*

Andrew Knight (2000). Basics of MATLAB and Beyond. CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Plorida 33431

Ati, O. F., Stigter, C. J., Iguisi, O. E. & Afolayan, J. O., (2009). Profile of rainfall change and variability in the northern Nigeria, 1953–2002. *Res. Journal Environmental Earth Science.*1 (2), 58–63.

Ayoade, J. O. (1976). A preliminary study of the magnitude, frequency and distribution Of intense rainfall in Nigeria. *Hydrological sciences bulletin* 3. 9: 419-429.

Ayoade, J.O. (2004). An introduction to climatology to climatology for the edition Ibadan spectrum books limited.

Ayodele, O. O. (2011). Temperature and rainfall variability in Kano, Katsina and Sokoto State of Nigeria.

Babalola, P. (2014). Heavy Rain Sacks Minna, Residents Stay Indoors. Received from: https://theeagleonline.com.ng/heavy-rain-sacks-minna-residents-stay-indoors/

Bello, N. J. (1998). Evidence of climate change based on rainfall records in Nigeria's weather, 53(12), 412–418.

Bello, N.J., (1998). A study of evidence of climate change based on rainfall seasonality and their liability of rainfall climate period in Nigeria. *Sustaining Africa. 4, 30–32.*

British Broadcasting Corporation Africa News (2018). Why does Nigeria keep flooding? Received from: https://www.bbc.com/news/world-africa-45599262

Brownlee, J. (2020). Four types of classification tasks in machine learning. Received from: www.machinelearningmastery.com

Bukka, U. A., Muhammad, B. & Yahaya, T. I. (2017). Effect of flooding on livelihood of communities in Muwo district, Mokwa Local Government Area, Niger State, Nigeria. *International Journal of Scientific & Engineering Research, 8(1615), 2229-5518*

Daramola1, M. T., Eresanya, E. O., & Erhabor, S, C. (2017). Analysis of rainfall and temperature over climatic zones in Nigeria. *Journal of Geography, Environment and Earth Science International. 2454-7352*

Douglas, M. M., Bunn, S. E. & Davies, P. M. (2005). River and wetland food webs in Australia's wet-dry tropics: general principles and implications for management. Marine and freshwater research, 56(3), 329-342.

Etuonovbe, A.K (2011). The devastating effects of flooding in Nigeria. An unpublished online Article retrieved on the 10/4/2017 via: http:11www.Fig.net/pub/fig2011/papers/tso6s/tso6/etuonovbe5002

Geetha. A, & Nasira, G. M. (2014). Artificial neural networks' application in weather forecasting. *International Journal of Computational Intelligence and Informatics, Vol. 4: No. 3.*

Gugulethu Z. N. (2013). Use of traditional weather/climate knowledge by farmers in the South-Western free state of South Africa: *Agrometeorological Learning by Scientists.* 4(4), 383-410. https://doi.org/10.3390/atmos4040383.

Han J, Kamber M, & Pei J. (2011). Data mining: concepts and techniques. New York: Elsevier; 2011

https://www.guru99.com/confusion-matrix-machine-learning example.html#:~:text=Example%20of%20Confusion%20Matrix%3A,-Confusion%20Matrix%20is&text=Below%20given%20is%20an%20example,world%20cup%2C%20and%20it%20won.

https://medium.com/analytics-vidhya/how-to-calculate-confusion-matrix-manually-14292c802f52

https://www.google.com.ng/amp/s/www.dataschool.io/simple-guide-to-confusion-matrix-terminology/amp/

Ibrahim, A. G. (2018). Postgraduate Lecture notes on the Physics of Lower Atmosphere. Physics dept. federal University of Technology, Minna

Julie M. D., & Kannan B., (2010). Significance of Classification Techniques in Prediction of Learning Disabilities. Cochin- 683 107, India.

Ifabiyi, I. & Ashaolu, P. (2013). Time series analyses of mean monthly rainfall for drought management in Sokoto, Nigeria. *Ethiopian journal of environmental studies and management, 6(5)*

Kowal, J. M. & Kanabe, D. T. (1972). An agro-climatic atlas of the northern states of Nigeria. Ahmadu Bello University Press, ABU, Zaria.

Kowal, J. M. and Kassam, A. H. (1978) Agricultural ecology of the savanna: A Study of West Africa. Clarendon Press, Oxford.

Kumar, S. (2020). Data Splitting Technique to fit any Machine Learning Model. Received from: www.towardsdatascience.com

Ladds, M. (2017). How much do disasters cost? A disaster loss from natural hazards in Australia. *International Journal of DRR. 10.106/j.ijdrr.2017.01.04*

Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. Copyright C John Wiley & Sons, Inc. ISBN 0-471-66657-2

Mitch, H. (1998). Tropical Rainfall Measuring Mission. Received from: www.earthobservatory.nasa.gov/features/TRMM

Nayak, D. R., (2013). A Survey on rainfall prediction using artificial neural network, *International Journal of Computer Applications (0975 – 8887), 72(16), 32–40.*

Ngamdu, M. B. (2015). An assessment of rainfall in Kumodugu Yobe river basin, Nigeria. *International Journal in Physical & Applied Sciences (Impact Factor) 2(865)*

NIMET (2018). An Overview of the Seasonal Rainfall Prediction. Nigerian Meteorological Agency (NIMET). Receieved from: http://www.nimet.gov.ng

NSEMA (2012). (https://www.vanguardngr.com/2012/10/rage-of-nature-flood-ravages communities-across-nigeria/) Received from www.https://www.vanguardngr.com

NSEMA (2018). (https://www.vanguardngr.com/2018/07/over-60-houses-affected-by-flood-in-niger-nsema). Received from www. https://www.vanguardngr.com

Ogungbenro, S. B. & Morakinyo, T. E. (2014). Rainfall distribution and change detection across climatic zones in Nigeria. Weather and Climate Extremes.

Oguntunde, P. G. (2011). Trends and Variability in Pan Evaporation and other Climatic Variables at Ibadan, Nigeria, Meteorological Applications, 19(4), 464–472.

Olaiya, F. & Adeyemo, A. (2013). Application of data mining techniques in weather prediction and climate change studies, *International Journal of Information Engineering and Electronic Business, 51-59*

Ologunorisa, T. E. & Tersoo, T. (2006). The changing rainfall pattern and its implication for flood frequency in Makurdi, Northern Nigeria. *Applied Science and Environmental Management, 10(3), 97-102*.

Oluleye, A., (2009). Change detection in rainfall anomalies across climatic zones in Nigeria. *Journal Meteorological Climate Scientist, 7: 6–10.*

Omonona B. T & Akintunde O. (2009). Rainfall effects on water use and yield of cocoa in Nigeria. *Continental Journal Agricultural Economics 3: 52 - 60,*

Oyewole, J.A, Thompson, A. M, Akinpelu, J.A. & Jegede O. (2014). Variation of rainfall and humidity in Nigeria. *Journal of Environment and Earth Science, 2224-3216*

Park, Y. S. & Lek, S. (2016). Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. Development in Environmental Modelling. *(28), 123-140.*

Pepin, N. C. (2017). A comparison of simultaneous temperature and humidity observations from the SW and NE slopes of Kilimanjaro: The role of slope aspect and differential land-cover in controlling mountain climate. Global and planetary Change. *(157), 244-258*

Pyle, D., (1999). Data Preparation for Data Mining. Received from: https://www.temida.si/~bojan/MPS/materials/Data_preparation_for_data_mining.pdf

Sarker, I. H., Kayes, A. S. & Watters, P. (2019). Effectiveness analysis of machine learning classification model for predicting personalised contex-aware smartphone usage. *Journal of Big data, 6(57). Doi.org/10.1186/s$0537-019-0219-y*

Theerthagiri, P., Jeena, J., Usha, A, R., & Vamsidhar., (2020). Prediction of COVID-19 possibilities using KNN classification algorithm. Received from: www. Researchsquare.com/articl/ rs-70985/v2 doi:10.21203/rs.3.rs-70985/v2

Tyagi, N. & Kumar, A. (2017). Comparative analysis of backpropagation and RBF neural network on monthly rainfall prediction, *International Conference of Invention Computer Technology. ICICT vol. 1*

Umar, A. T. (2012). Analysis of extreme rainfall events and risk of drought and flood occurrences in Nigeria. *Nigeria Geographical Journal, Volume 8(2).*

Weather Parameters (2018). Received from: (http://www.nationalgeographic.org/encyclopedia/weather/).

Zhang, S. & Zhou, H., (2016). Short-Term Water Level Prediction using Different Artificial Intelligent Models, *in 5th International Conference on Agro-Geoinformatics, Agro Geoinformatics.*