DSpace Institution

DSpace Repository **http://dspace.org**

Electrical Engineering Thesis

2021-06

PREDICTION OF MALARIA ABUNDANCE USING ARTIFICIAL NEURAL NETWORK (ANN) IN ETHIOPIA: - A CASE STUDY OF AMHARA REGION

ASCHALEW, ABEBE BELETE

http://ir.bdu.edu.et/handle/123456789/12681

Downloaded from DSpace Repository, DSpace Institution's institutional repository

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

PREDICTION OF MALARIA ABUNDANCE USING ARTIFICIAL NEURAL NETWORK (ANN) IN ETHIOPIA: - A CASE STUDY OF AMHARA REGION

BY

ASCHALEW ABEBE BELETE

BAHIR DAR, ETHIOPIA

JUNE 2021

PREDICTION OF MALARIA ABUNDANCE USING ARTIFICIAL NEURAL NETWORK (ANN) IN ETHIOPIA: - A CASE STUDY OF AMHARA REGION

BY

ASCHALEW ABEBE BELETE

A Thesis submitted to the School of Research and Graduate Studies of Bahir Dar Institute of Technology, BDU in partial fulfillment of the requirements for the degree of

MSc. in the Computer Engineering in the Faculty of Electrical and computer Engineering

Advisor Name: Dr. Eng. Getachew Alemu

BAHIR DAR, ETHIOPIA

July 17, 2021

Declaration

This is to certify that the thesis entitled "**PREDICTION OF MALARIA ABUNDANCE USING ARTIFICIAL NEURAL NETWORK (ANN) IN ETHIOPIA: - A CASE STUDY OF AMHARA REGION**", submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical and computer Engineering under Faculty of computer Engineering ,Bahir Dar Institute of Technology , is a record of original work carried out by me and has never been submitted to this or any other institution to get any other degree or certificates. The assistance and help I received during the course of this investigation have been duly acknowledged.

Aschalew Abebe		9/11/2015
Name of the candidate	signature	Date

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND GRADUATE STUDIES

FACULTY OF ELECTRICAL AND COMPUTER ENGEENERING

Approval of thesis for defense result

I hereby confirm that the changes required by the examiners have been carried out and mincorporated in the final thesis.

Name of Student: **Aschalew Abebe** Signature _____ Date 9/11/2013

As members of the board of examiners, we examined this thesis entitled "**Prediction of Malaria Abundance using Artificial Neural Network in Ethiopia : - A case study of Amhara Region**

"by **Aschalew Abebe**. We hereby certify that the thesis is accepted for fulfilling the requirements for the award of the degree of Masters of Science in "Computer Engineering".

Board of Examiners

Name of Adviser	Signature	Date
Getachew Alemu (Ph.D)		9/11/2013
Name of External examiner	Signature	Date
AbebeTesfahune (Ph.D.)		9/11/2013
Name of Internal Examiner	Signature	Date
Adissalem Gemta (Ph.D.)		9/11/2013
Name of Chairperson	Signature	Date
Molla Atanaw		9/11/2013
Name of Chair Holder	Signature	Date
Molla Atanaw		9/11/2013
Name of Faculty Dean	Signature	Date
Tewodrose Gera		9/11/2013

© 2021
ASCHALEW ABEBE BELTE
ALL RIGHTS RESERVED

To God and His Mother

ACKNOWLEDGMENTS

I would like to express special thanks to my Advisor Dr. Getachew Alemu, for sharing his knowledge and experience for setting a clear example of excellence in research. I would like to thank his for encouraging my research and for allowing me to grow my idea in Research, I am also very grateful for ECE faculty staffs for those continue support. Finally, to my dear family, many thanks for everything. All are done by GOD then, I thank my God, Thanks Marry i will keep on trusting you for my future. Thank you, Lord.

ABSTRACT

Malaria affects more than 110 Million peoples worldwide with approximately 2414 diseases a day in average each year. Among this majority of the disease occur in Africa especially in sub Saharan including Ethiopia. More than 75% of the total area of Ethiopia is malarious, making malaria is the leading public health problem in Ethiopia. Neural Network was applied in a great variety of areas for data mining and knowledge discovery, Prediction, systems modeling, optimization and pattern recognition. Study has also shown the close relation between malaria incidences, weather pattern and clinical History.

In these work we are implemented the ML algorithm to predict Malaria cases in Amhara region using clinical and environmental data. Those back history data are available from Ethiopian Meteorology Agency, Amhara Public health institute and various health centers of Amhara Region for train and predict the model. The collected data are used for input and output to the model for create awareness and alarm about malaria disease including transmission and their reproductive season using python simulation tool.

The research singled out LSTM as the basis for the model, after considering various algorithms and their suitability for different problem domains . Four input and one output are used for deploy the model. The dataset are prepared using twelve years of data that are splitted in to 8 years of data used for traning,2 years of data used for validating and 2 years of data are used for testing. That means the model used approximately 70% data used for training, 15% for validating and 15% for testing. The performance of LSTM is shown to be the best fit model with minimum Mean Squared Error (MSE) of 0.096054 in amhara region, 0.049267 in south Gondar and 0.067914 in west Go jam which is small enough, and it indicates that the predictive model performs well.

Keyword: LSTM; prediction; neural network; Malaria

TABLE OF CONTENTS

Contents

DECLARATION i

ACKNOWLEDGMENTS v

ABSTRACT vi

LIST OF FIGURE……xi

LIST OF TABLE.....Xiii

LIST OF ABBREVIATIONS xiv

CHAPTER ONE..... 1

1. INTRODUCTION 1

1.1 BACKGROUND 1

1.1.1 Socio-Economic Impacts of Malaria in Ethiopia. 2

1.1.2 Effect of Climate Change on the Prevalence of Malaria in Ethiopia: 3

1.1.3 Overview of climatic factor for Malaria disease in Amhara Region 3

1.1.4 Prevention and control of malaria 4

1.2 Statement of the problem……..……4

1.3 Objective of the study…….....5

1.3.1 General objective ..... …..5

1.3.2 Specific objectives..... 5

1.4 Research Methodology 6

1.5 Scope of the study 6

1.6 Significance of the Study 6

1.7 Organization of the thesis..... 7

CHAPTERTWO ……..... 8

2. Review of Literature..... 8

2.1. Malaria Distribution in Ethiopia..... 8

2.2. Malaria in Amhara Region, Ethiopia..... 9

2.3. Machine learning techniques for Malaria prediction..... 10

2.4. Special selected areas of Amhara Region for the study 10

2.4.1. Artificial Neural Network..... 10

2.4.2 Related Works 15

CHAPTER THREE 18

3. METHODOLOGY 18

3.1. Introduction..... 18

3.2. Study area and Data Collection 19

3.2.1. Study Area..... 19

3.2.2. Data Collection.....20

3.3. Data Preprocessing..... 26

3.1.2. Missing Values handling…..... 26

3.1.2. Normalization 28

3.4. Model Development..... 33

3.4.1. Recurrent Neural Networks……..…….....33

3.4.2.The Problem of Long-Term Dependencies……..…...34

3.4.3. LSTM Networks and their Architecture…….....35

3.4.4. LSTM working principles ……..……...38

3.5. Setup the dataset..... 41

3.6. Error metrics..... 44

3.7. Model Implementation..... 45

CHAPTER FOUR..... 46

4. RESULTS AND DISCUSSION..... 46

4.1.Evaluating the model..... 46

4.2.Result..... 46

4.3.Model Diagnostic Plots..... 59

4.3.1. LSTM Under fitted Model Losses..... 59

4.3.2. LSTM Good fitted Model Losses……..……...60

4.3.3. LSTM Over fitted Model Losses ..... …..61

4.4. Discussion..... 62

CHAPTER FIVE 63

5. CONCLUSION AND RECOMMENDATION 63

5.1 Conclusion 63

5.2 Recommendations for Further Research..... 64

REFERENCES..... 65

APPENDIXES..... 68

Appendix 1: South Gondar Monthly Malaria Cassese (2006-2017) 68

Appendix 2: West Go jam Monthly Malaria Cases (2006-2017).....68

Appendix 3: West Go jam Monthly T max(2006-2017) 69

Appendix 4: West Go jam Monthly T min(2006-2017) 72

Appendix 5: West Go jam Monthly RH (2006-2017)..... 70

Appendix 6: West Go jam Monthly RR (2006-2017) 70

Appendix 7: South Gondar Monthly Humidity (2006-2017) 71

Appendix 8: South Gondar Monthly RR (2006-2017)..... 71

Appendix 9: South Gondar Monthly T max (2006-2017)..... 72

Appendix 10: South Gondar Monthly T min (2006-2017).....72

Appendix 11: South Gondar Dataset all Attribute in one (2006-2017) 76

Appendix 12: West Go jam Dataset all Attribute in one (2006-2017)..... 80

Appendix 13: South Gondar Normalized Dataset all Attribute in one (2006-2017) 84

Appendix 14: West Go jam Normalized Dataset all Attribute in one (2006-2017).....89

LIST OF FIGURES

Figure 1.1: Number of clinical malaria cases reported in 2001 by age group and Pregnancy Status..... 2
Figure 2.1: Malaria case trend by year in Amhara region, 2000 – 2010 EFY……..…8
Figure 2.2: The basic functionality of a simple artificial neurons..... 12
Figure 2.3: Depicting basic architecture of a two-layer FFNN. 13
Figure 3.1: System flowchart..... 18
Figure 3.2: Study area location 19
Figure 3.3: Specific study location..... 20
Figure 3.4: Amhara Region Total Malarial Cases in chart (2006-2017) 22
Figure 3.5: Amhara Region Monthly Maximum Temperature in chart (2006-2017) 23
Figure 3.6: Amhara Region Minimum Temperature in chart (2006-2017)..... 24
Figure 3.7: Amhara Region Monthly Relative Humidity in chart (2006-2010)..... 25
Figure 3.8: Amhara Region Rainfall in chart (2006-2017)..... 26
Figure 3.9: Amhara Region Normalized Minimum Temperature in chart (2006-2017) 29
Figure 3.10: Amhara Region Normalized RR in chart (2006-2017) 30
Figure 3.11: Amhara Region Normalized RH in chart (2006-2017)..... 31
Figure 3.12: Amhara Region Normalized Maximum Temperature in chart (2006-2017) 31
Figure 3.13: Amhara Region Normalized Malaria cases in chart (2006-2017) 32
Figure 3.14: Amhara Region Normalized Malaria cases in chart (2006-2017 33
Figure 3.15: Looped Recurrent Neural Networks 33
Figure 3.16: An unrolled recurrent neural network 34
Figure 3.17: The repeating module in a standard RNN contains a single layer 36
Figure 3.18: The repeating module in an LSTM contains four interacting layers. 37
Figure 3.19: Typical LSTM network.....38
Figure 3.20: Forgate gate of LSTM cell 39
Figure 3.21: Input gate of LSTM networks 40
Figure 3.22: Output gate of LSTM network 40
Figure 3.23: Amhara region monthly validating output dataset…….....……..…42

Figure 3.24: Amhara region monthly validating output dataset……..43

Figure 3.25: Amhara region monthly testing input dataset……43

Figure 3.26: Amhara region % of < 5-year Malaria from total Malaria by 2009 and 2010 EFY……44

Figure 4.1: Actual and Predicted malaria of the year 2016 using graph in Amhara Region 47

Figure 4.2: Actual and Predicted malaria of the year 2016 using Histogram in Amhara Region 47

Figure 4.3: Actual and predicted malaria of the year 2017 using graph in Amhara Region 48

Figure 4.4: Actual and predicted malaria of the year 2017 using histogram in Amhara Region 49

Figure 4.5: Actual and predicted malaria of the testing data using graph in Amhara Region 49

Figure 4.6: Actual and predicted malaria of the testing data using Histogram in Amhara Region.50

Figure 4.7: Model Loss of the developed model for Amhara Region 50

Figure 4.8: Actual and predicted malaria of the year 2016 using graph in S/Gondar 51

Figure 4.9: Actual and predicted malaria of the year 2016 using Histogram in S/Gondar 52

Figure 4.10: Actual and predicted malaria of the year 2017 using graph in S/Gondar 52

Figure 4.11: Actual and predicted malaria of the year 2017 using Histogram in S/Gondar 53

Figure 4.12: Actual and predicted malaria of the testing data using graph in S/Gondar..... 53

Figure 4.13: Actual and predicted malaria of the testing data using Histogram in S/Gondar 54

Figure 4.14: Model Loss of the developed model for S/Gondar Region 54

Figure 4.15: Actual and predicted malaria of the year 2016 using graph in W/Gojam 55

Figure 4.16: Actual and predicted malaria of the year 2016 in W/Gojam Using Histogram 56

Figure 4.17: Actual and predicted malaria of the year 2017 using graph in W/Gojam 56

Figure 4.18: Actual and predicted malaria of the year 2017 in W/Gojam Using Histogram 57

Figure 4.19: Actual and predicted malaria of the testing data using graph in W/Gojam 57

Figure 4.20: Actual and predicted malaria of the testing data using Histogram in W/Gojam 58

Figure 4.21: Model Loss of the developed model for W/Gojam Zone 58

Figure 4.22: Representation of LSTM under fitted model Losses Graphically 59

Figure 4.23: Representation of LSTM Good fitted Model Losses Graphically..... 60

Figure 4.24: Representation of LSTM Over fitted Model Losses Graphically 61

List of Table

Table 3.1: Amhara Region Monthly Malarial cases (2006-2017)..... 21
Table 3.2: Amhara Region Monthly maximum temperature (2006-2017) 22
Table 3.3: Amhara Region Relative humidity (2006-2017)..... 23
Table 3.4: Amhara Region Monthly Rainfall (2006-2017)..... 24
Table 3.5: Amhara Region Monthly Monthly minimum Temperature (2006-2017)..... 25
Table 3.6: Amhara Region Normalized Minimum temperature (2006-2017) 28
Table 3.7: Amhara Region Normalized Relative Rainfall (2006-2017) 29
Table 3.8: Amhara Region Normalized RH (2006-2017) 30
Table 3.9: Amhara Region Normalized Malarial cases (2006-2017 32
Table 3.10: Amhara Region Training dataset(2006-2007) 41
Table 3.11: Amhara Region validating dataset (2014-2015)..... 42
Table 4.1: Amhara Region Actual and predicted Malaria cases of 2016..... 46
Table 4.2: Amhara Region actual and predicted Malaria of 2017 48
Table 4.3: Actual and predicted Malaria of 2016 in W/Go jam and S/Gondar Zone 51
Table 4.4: Actual and predicted Malaria of 2017 in W/Go jam and S/Gondar zone..... 55

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BDU	Bahir Dar University
BIT	Bahir Dar Institute of Technology
CNN	Convolutional Neural Network
DVBD	Department OF Vector-Borne diseases
FF	Feed Forward
FFNN	Feed Forward Neural Network
GMDH	Group Method of Data Handling
LSTM	Long-Short Term Memory
MLP	Multi-Layer perceptron
MLR	Multi-Layer Regression
NN	Neural Network
PHC	Primary Health Center
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
ROC	Region of Convergence
S/Gondar	South Gondar
SVM	Support Vector Machine
TNR	True Negative Rate
W/Go jam	West Go jam

CHAPTER ONE

1. INTRODUCTION

Approximately up to 2 million people die by malaria around the world annually, mostly in Africa and half of them are children (T. Deress,2019). Malaria is one of the main problems of Sub-Saharan Africa, especially in Ethiopia. The majority of the people in Ethiopia are not aware of malaria and their transmission factors. It is dependent on weather and climate factor to reproduce and survive their life. Children including newborns and infants that are less than one years of age are the most vulnerable groups. From 75% of the total landmass of Ethiopia is malarious and 60% of the population is at risk of the disease (D. G. Ayele, T. T. Zewotir, and H. G. Mwambi,2012). The Report of Federal Ministry of Health annual performance indicated that between September 2014 and August 2015 total of 2,174,707 malaria cases and 662 deaths were documented with a case fatality rate of 0.03%. Among these the majority is in Amhara region compared with another region, which is 610,486 cases are reported followed by Oromia which is 430,969 malaria case and Southern Nations, Nationalities and Peoples Region (SNNPR) which is 375,746 cases are reported (D. G. Ayele, T. T. Zewotir, and H. G. Mwambi,2012). To predict malaria abundance, different Mechanism models have been developed. The malaria prediction model starts from early works of Christopher in 1911 (S. Thakur and R. Dharavath,2018), till today, many models have been developed. Each work has a different interpretation and verity of prediction in their own mechanism. Those methods are used for prediction in a certain period of time, but in previous work, the data like clinical and environmental data for best neural network model in Ethiopia is not explored.

.

1.1 BACKGROUND

The problem of malaria is very severe in Ethiopia where it has been the major cause of illness and death for many years. According to records from the Ethiopian Federal Ministry of Health more than 50 million people are at risk from malaria and four to five million people are affected by malaria annually (APHI, 2010). We will conduct our study in the specific geographical location of Amhara region for prediction of malaria abundance. Machine learnings have so many applications that involve computers learning from data (E. Silva and J. Cesar, 2013). For simple tasks assigned to computers, it is possible to program

algorithms telling the machine how to execute all steps required to solve the problem at hand. More advanced tasks can be challenging for a human to manually create the needed algorithms.

ML are more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

1.1.1 Socio-Economic Impacts of Malaria in Ethiopia.

An estimated 300 million cases of acute malaria occur each year globally, 90 percent of the 2 million malaria deaths occur in Africa, mostly in young children and especially in sub-Saharan Africa including Ethiopia (H. S. Taffese *et al.*,2018). The malaria season coincides to a peak economic activity in rural Ethiopia when both rainfall levels and temperature are high and conducive for the growing of subsistence crops. An optimal climatic regime for socioeconomic activities in rural Ethiopia also favors the reproduction, propagation, and there by the preying up human blood of vector mosquitoes.

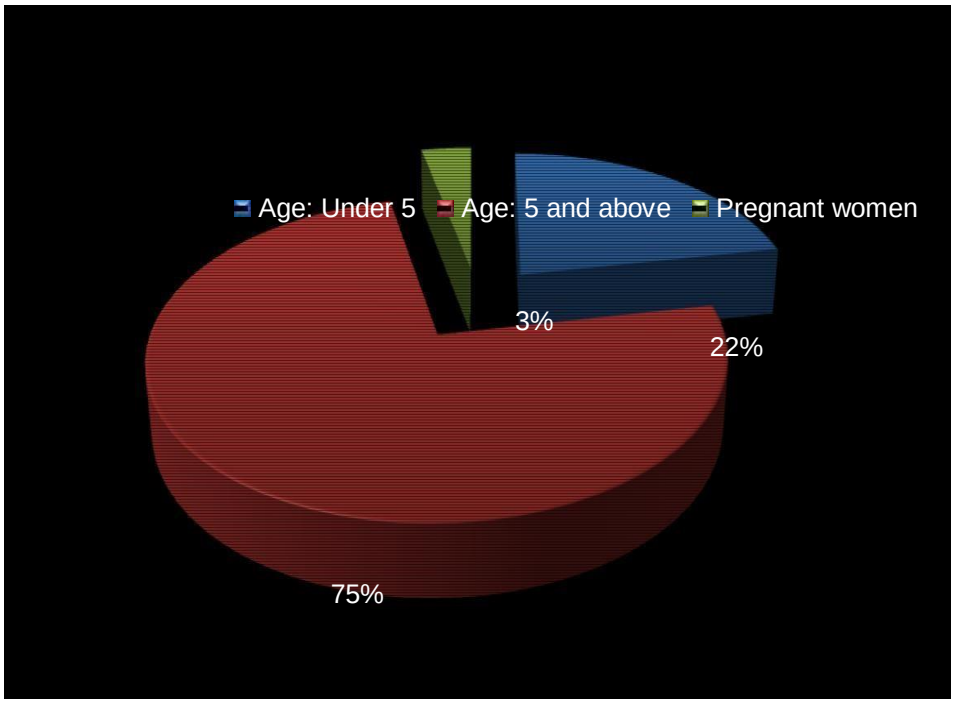


Figure 1.1: Number of clinical malaria cases reported in 2001 by age group and Pregnancy Status.

The timely detection of malaria is challenged and difficult by inconsistent commitment in passive surveillance across Ethiopia and their geographic region that lack of resources for adequate data collection.

1.1.2 Effect of Climate Change on the Prevalence of Malaria in Ethiopia:

When temperature increases and a change in the level of precipitation shows that, there are some aspects of climate change that may affect vector-borne diseases such as malaria. The optimum hot and cold environments are good places for mosquitoes to breed (M. Sewe, J. Rocklöv, J. Williamson, M. Hamel, and A. Nyaguara,2015). Climate are an important contribution in the seasonal pattern, that are carried and transmitted for vectors, because the vector animals often survive in particular climate conditions. The mosquitoes breeding happen in specific weather conditions to be a species that can transmit a disease and, if there is an infected population in the region the disease is more spread in that area. Malaria is the most parasitic disease in the world (M. Taye, F. Zewdu, and D. Ayalew, 2013).

1.1.3 Overview of climatic factor for Malaria disease in a geographical location of Amhara Region

Malaria are mosquito-borne disease globally that are directly dependent on climatic factor (E. Loha and T. M. Lunde, 2010). Anopheles Mosquito are their own optimum temperature to leave and reproduce.

(Rahila Parveen†, Akhtar Hussain Jalban.et.al, 2017). Very high temperatures and very low temperatures including rainfall and humidity are unable to comfort for the vector. The spatial occurrence of mosquitos are largely determined by climatic variables. Malaria occurs where there is sufficient perception to establish suitable mosquito habitat for vector species. An adequate humidity to enable highly survival of mosquitoes, elevated temperatures, optimum wind speed, and rainfall to sustain quick gonotrophic and saprogenic cycles. So temperature determines the elevation of a limit of stable malaria transmission.

The temperature, humidity, rainfall and wind speed in the region varies from month to month and area to area annually. These attributes are different from area to area, not only for Amhara region but also in Ethiopia there are three weather conditions. Dega, kola and Waina Daga, or hot zone, cold zone, and medium zone are the weather condition that accour in this region. In this Region the cold zone are Shewa, Dessie, Deberatabour, Nefase mawecha, Giant, Estie, Fareta, Debera Markose, semada which is mean annual temperature is below 10-15°C. In these areas the mosquito is unable to comfort to lay there egg

and difficult to survive and reproduce, so the malaria distribution in this zone are low. Hot zone geographical area have mean annual temperature varies from 20-25°C. Bahir Dar, Woreta, Fogera, Gondar, Metema, Kombolcha are some hot geographical location of Amhara region. These temporal zone mosquito is suitable to reproduce and survive, then the percentage of malaria is increase in this zone, but it depends on annual weather condition. It is different from month to month.

The majority of malaria transmission are from February to the end of June and September to December depends on the start and end of summer rainfall (A. Arifianto, 2014). However, in July and August, there are high rainfall and the weather are very cold then the mosquito eggs are clear by floods. Mosquito are unable to comfort to reproductive and survive at this time, the malaria transmission is of course low.

1.1.4 Prevention and control of malaria

Malaria are prevented by a goober and DDT (H. S. Taffese *et al.*,2018), but in Ethiopia there are so many limitation to supply these for the rural area. The majority of the people are unawareness about this prevention, not only these but also 85% of the people are farmers that are unwariness about the prevention and control of malaria, including transmission factors. So to control the spread of malaria those people must aware about malaria transmission and the vector breeding factors like humidity, rainfall, wind speed and temperature.

1.2 Statement of the problem

Malaria is one of the main problems in the world especially developing country like in sub-Saharan Africa including Ethiopia. In Ethiopia majority of the people lives in ruler area almost 85% population are farmer, not civilized and not idea about the diseases and their transmission of malaria. AI methods are used for analyze malaria by related with climatic factor and gain intelligence, that use this intelligence in prediction that can be used to model such a tool. In Ethiopia there is no efficient disease prediction model including malaria. These work designed and implemented prediction of malaria diseases using RNN which are a type of ANN to create awareness that helps for reducing the number of malaria diseases.

The thesis will implement and design the prediction model using advanced technological Model known as Machine learning. The weather factor are used as an input for predictive analysis in the amhara region. Prediction is important to create awareness about Mosquito and malaria disease by gather past patient record from Hospital, primary health center in addition to Environmental data. **(**Ivan Kiskin1;2, Bernardo P_erez Orozco1;2, et.al.(2017) touched this area when they applied using NN model, However, the proposed models are statistical and traditional machine learning. Hence, the study aims to explore the predictability of Malaria and efficient data analysis as well as prediction model using deep learning algorithms for amhara region. Generally, the research issues are formulated as follows: ❖ Which deep learning model is best fit in prediction of Malaria abundance in amhara region?

- ❖ Which climate features are appropriate for mosquito and malaria prediction and how to relate them?
- ❖ How the disease analysis approach is determinant for Amhara Region prediction?

1.3 Objectives of the study

1.3.1 General objective

❖ Design and Implement Artificial Neural Network Model for prediction of Malaria abundance in Amhara Region.

1.3.2 Specific objectives

- ❖ Create awareness about Malaria disease, Mosquito reproductive area and season that causes malaria in Ethiopia (Specific geographical location of Amhara)
- ❖ Investigate the prediction of malaria diseases by relate with weather attribute ❖ Design NN model which best fits to the dimension and nature of collected data ❖ Train and test the developed model.

1.4 Research Methodology

The procedures included in this Thesis are:

Literature Review: A number of published ideas about Malaria Abundance, prediction, and the relation of climatic factors with malaria in both AI and statically in journals, articles, lectures, materials have been reviewed.

Modeling: Model AI algorithm using Python Anaconda software, finally Jupiter not book simulation software tools are used for training, test, predict the AI, model.

1.5 Scope of the study

The proposed system will able to predict malarial cases for the Amhara region using the station data. In this thesis, 12 years of data are used for training and prediction. The prediction focuses on all seasons of Ethiopia and used only three meteorological parameters as input and the malarial cases as an output.

1.6 Significance of the Study

To the scientific world:

- Provides LSTM model is good for time-series data ❖ Show initial points for new findings for other researchers To the community:
- Minimize the impact of malarial disease by creating awareness
- Improve health status of the people by avoiding the malarial disease
- Save money that extravagance for clinical use to increase productivity

1.7 Organization of the thesis

This Thesis has a total number of five chapters and the discussion of the study is organized as follow:

Chapter One: Provides a brief description of the background, the disease distribution, AI, problem statement, objectives, research methodology, scope and significance of the thesis.

Chapter Two: Discusses literature reviews on Malaria and different insects that affect and herm the human that reproduce in different weather condition, and also Machine learning methods are applied for these work proposed by different researchers. Besides, the common gaps of the reviewed works and the way how we fill in the gaps are described.

In chapter three: Discusses the methodologies followed to conduct this thesis, which comprises problem formulation, data collection, data analysis and also the detailed description of the proposed LSTM model is discussed, and the description of the system, which is composed of three parts (training, validation, and testing) are presented. The components that compose the system (data collection, preprocessing, normalization, setup the dataset), and methods are described in detail.

In chapter four: Reveals the results obtained during the simulation and the discussion about the results found. Showing the results of the proposed model in the graph and the discussion of the result described briefly.

In chapter five: we will summarize the major finding of this research work, and also the recommendation at some points will be outlined for future work

CHAPTER TWO

2. Review of Literature

2.1. Malaria Distribution in Ethiopia

The distribution of malaria in Ethiopia varies from place to place due to the above factors directly or indirectly affecting the pattern of malaria transmission. The exact number of people that are sick and dying by malaria every year in Ethiopia is not exactly known. However, it is known that millions of people get sick and tens of thousands of people die due to malaria every year (D. R. Tefera, S. O. Sinkie, and D. W. Daka, 2020).

Distribution of malaria is not uniform there are an area where the risk of malaria is high and there are an area where the risk is low. Malaria transmission peaks annually from September to December and April to May. The Figure given below is based on confirmed malaria cases that have come to the attention of health workers and institutions, the yearly bar graphs are just surrogates for overall infection levels estimated at about 5 million a year that shows the incidence of malaria in Ethiopia is not only high but also rising. The overall trend is one of a gradual progression from around four-fifths of a million confirmed cases a year in 2000 to almost a million and a half in the year 2000-2010 reporting period, this represents an increase of about 80 percent (D. R. Tefera, S. O. Sinkie, and D. W. Daka, 2020). The turn of the century 2000 to 2010 marked the worst decade of malarial infections in recent history with the number of confirmed cases exceeding one and a half million a year.

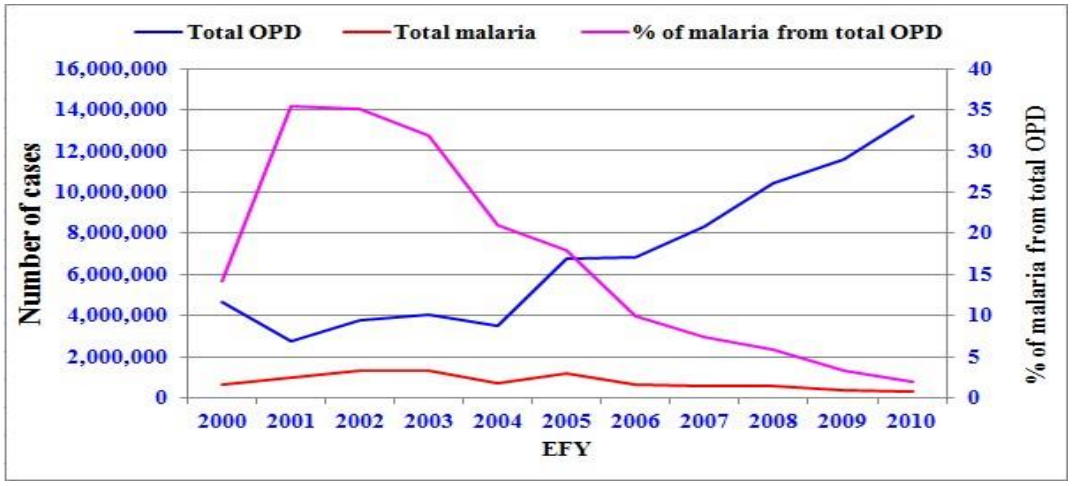


Figure 2.1: Malaria case trend by year in Amhara region, 2000 – 2010 EFY

2.2 Malaria in Amhara Region, Ethiopia

Amhara's drastic altitudinal range from 506 meters at the Blue Nile Gorge, to 4533 meters at Ras Dashen, makes the prediction of the timing and geographic range of malaria outbreaks particularly challenging. By 2012 GC the Amhara Region counted 1,127,241 malaria cases, out of a population of 19,867,817 inhabitants (D. Lankir, S. Solomon, and A. Gize,2020). The malaria epidemics in nonendemic regions of the country of the Amhara Region largely occur as a result of above-average temperatures and precipitation (A. Alelign, Z. Tekeste, and B. Petros, 2018). Precipitation is strongly linked to the prolonged occurrence of water bodies for female mosquitoes to lay their eggs and for successful larval development; these egg-laying sites are most common in deforested regions, like agricultural fields. The temperature are impacts on the length of larval development, mosquito survival, and parasite development cycle. The elevated temperatures increase the development rate of both mosquito larvae and the Plasmodium parasites (R. Parveen *et al.*,2017). Within increased development rates, a higher number of mosquito generations arise and is in a greater abundance. An increased malaria transmission rates with higher temperatures result from elevated feeding rates of the female adult mosquito that increases the likelihood of malaria transmission to humans. The effect of increased rainfall on malaria transmission and risks for malaria outbreaks are composed by increased temperatures and increased humidity.

A changing climate will influence the variables that influence malaria transmission like, temperature and precipitation, that change to altitudinal distribution of malaria in these region. Malaria is affecting the majority of the world start from the past to the present, because of these many works are investigating to minimize and protect from the disease, even though malaria vector directly depends on climatic factor. The use of ANN in acoustic detection, prediction, and classification of Species dates back to at least the beginning of the century, with the first approaches addressing the identification of bat echolocation calls (E. Gofer, March, 2014). Both manual, Medical and algorithmic techniques have subsequently been used to identify the vector and their disease from the past till now. This section will review the use machine learning approaches in bioacoustics, in particular with respect to insect recognition, transmission, prediction, and identification especially in vector born disease like malaria.

2.3 Special selected areas of Amhara Region for the study

In Amhara Region, South Gondar Zone and West Go jam Zone are used for additional malaria forecasting. Both Zones are high malaria distribution proportional to their weather factor. Their geographical construction are made up of more than ten worada and many kebeles. According to the data from Ethio mereological agency, the Zones are their own weather variability. The weather and malarial data used for these work are available at the appendixes part of these thesis

2.4 Machine learning techniques for Malaria prediction

Machine learning is a field of science, aiming specifically at learning and extracting knowledge from datasets, a model developing real-world simulations; apply prediction, classification, and pattern recognition methodologies on the input data. Time series analysis has essentially helped in the development of both traditional and intelligent methods. Traditional methods require assumptions about the characteristics of the data. The nature of Malarial and weather data are non-linear, Frequency, concentration and amount are the main characteristics of time series malarial forecasting. These values can be varied from one position on earth to another position of the earth and from one time to another time (shivhare, Rahul, Dwivedi,&Hindu,2019).

The most widely used methods for malarial forecasting are: statistical methods and numerical models. Every statistical model has some drawbacks. A combination of Auto Regression and moving average together form a general and useful class of the time series model known as Auto Regression Moving Average Model (R.K Agrawal, Ratnadip Adhikari). The hybrid models are only useful for stationary time series data. The statistical pattern cannot identify nonlinear patterns and irregular trends in the time series. Among the several machine-learning models in the sub-field of data prediction, neural network techniques are valid and useful alternatives to statistical methods in solving nonlinear data problems (A. Gupta, 2016).

2.4.1 Artificial Neural Network

ANN is an interconnection of neural or neurons that works artificially, and an interconnection of neuron in a computing system inspired by the biological neural network to train and learn like animal brain. The main goal of ANN are to solve problems in the same way that a human brain would, and

have been used on a variety of tasks, like computer vision, speech recognition, Machine translation, and social network filtering, medical diagnosis and so on(M. P. Darji,2015).

In most cases, an ANN is an adaptive method that changes its organization based on external or internal information that flows through the network during the learning phase. Neural Networks depending on the architecture, provides association, classifications, clusters, prediction, and forecasting to the data mining industry(Mohini P. Darji,2015; Vipul K. Dabhi et al.,2015). Network means an interconnection of more than one topology or nodes to work and solve many problems and tasks. In these work network in ANN are an interconnected of more than one neuron by forming a topology to solve an artificial and related problem like a human brain, so ANN is a collection of interconnected units or nodes called artificial neurons. Networks consist of units or nodes and connections between them for a common job to solve a desired problem in place of human brain, in order to produce an output they take input and process it. The connection is used for information flow by input data, process it to get the desired output can be directed. If the connection are directed there is a constraint on which only lets the information flow in one or the other direction.

A node can be different in complexity and function to combine the power of several nodes to be able to complete tasks that are more difficult. Artificial neural network is one type of network, which consists of artificial neurons that can be connected in a different way in order to obtain the desired complexity and functionality. The complexity of ANN considerably abstracted from biological neural networks are based on the idea of neural neuron send and receive signals through synapses. Artificial neurons are the basic building block of ANN, designed and characteristics are derived from the observation of natural neurons in biological neural networks. As an overview of the structure and functionality of artificial neurons with its inputs, weights, transfer, activation function, threshold and output including general structural and working principles of artificial neurons as follows in the figure ((shivhare, Rahul, Dwivedi,&Hindu,2019) .

An artificial neuron can be fed with multiple number of neuron input, then depending on the weight value the total input influence to the neuron will differ as well as the computation of the transfer function and outputs. The higher weight values increases the power and influences of the associated input (K. Y.

Lee, N. Chung, and S. Hwang, 2016) because of all the inputs are multiplied by their cross ponding weight and that weight will affect the neuron's output, as follows in Equation

netj= $\sum_{i=1}^{n} wijxi + b$ Where:

- j denotes an actual neuron number
- xi is the input value where i goes from 1 to n

The final output from the neuron is computed as $OJ = f(netj)$ where f is an activation function, − the activation function defines the properties of an artificial neuron and is chosen depending on the problem that the artificial neural network needs to solve. The most commonly used activation functions are variations of the sigmoid function.

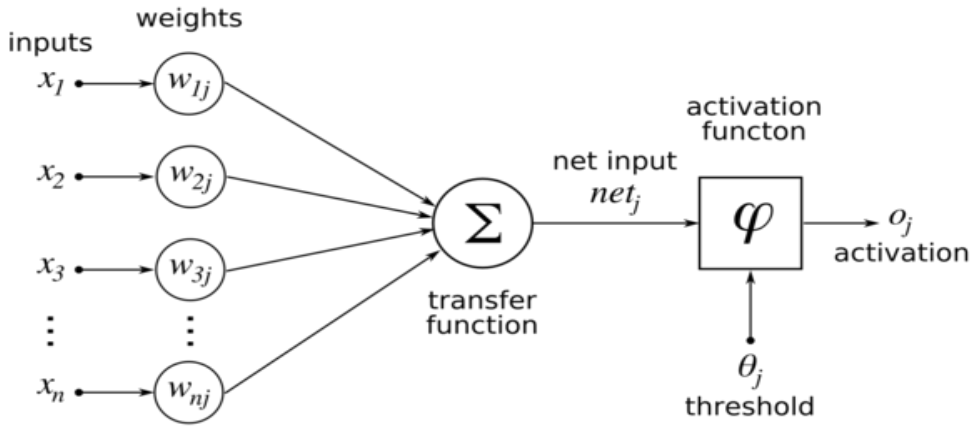


Figure 2.2: The basic functionality of a simple artificial neuron.

An artificial neural network with feed-forward topology is called a feed-forward neural network and is the most popular and widely used ANN topology. FFNN consist of layers of interconnected neurons where each neuron in the input layer is connected to all neurons in the next layer and each neuron in this layer is connected to all neurons in the next layer, and so on until finally they are connected to the final output layer. There is no limitations on the number of layers, but increasing the number of layers does increase the complexity of the network then the neurons are difficult to learn. All layers between the input and output layers are called the hidden layers, simply meaning that they are neither input nor output neurons (M. Obach, R. Wagner, H. Werner, and H. H. Schmidt, 2001).

The figure shown below illustrates a multilayer Feed forward neural network.

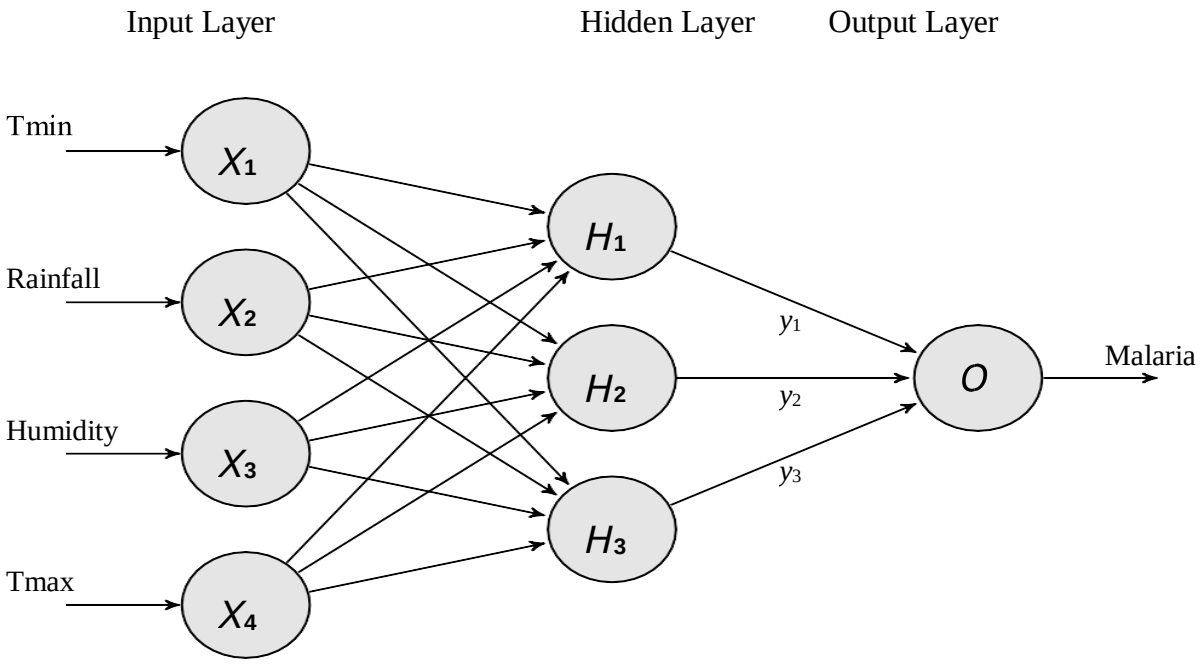


Figure 2.3: Depicting basic architecture of a two-layer FFNN.

The input layer is made up of neurons X_1 up to X_4 and the hidden layer are H_1 to H_3 in the network. The popularity of FFNNs derives from the fact that they have been successfully applied to the purpose of function approximation, which means they are able to approximate any continuous function arbitrarily well given the right number of neurons in the hidden layer. For this case Feedforward, neural networks are applicable in situations with unknown nonlinear relationships, making it suitable for the prediction purpose (A. Gupta, 2016).

The learning or teaching process of an artificial neural network is essential for its functionality; this is carried out by one out of two types of training algorithms based on different concepts. Supervised learning means that in a given vector of inputs to the network, it will produce the respective outputs and compare it to the correct answers at hand or the target outputs (G. Di Franco and M. Santurro, 2020). The main objective is to reduce error by comparing the learning and desired output. Error=Network output-Target, the answers are available to the control process and can be compared to the computed outputs and this is utilized in the learning process. Additionally, supervised learning is divided into methods using two other types of learning, the one used in this paper are error correction (A. Dey and A. S.,2016).

The magnitude of the error together with the input determines the correction to the weights. Finally, the goal is to minimize or fully eliminate this error.

To use an ANN it has to train and learn how to handle the input data. Those the learning process corresponds to setting the values of weights and biases, so it minimizes the error between the outputs and expected values, different training algorithms can be used for the Artificial Neural Networks that depends on the topology and activation function. The back propagation algorithm is used in FFNNs together with a sigmoid function because of its easily computed and continuous derivative (M. Selvam, B. Maniam, D. Kannaiah, C. Kathiravan, and T. Vadivel, 2018). Also, data is only fed forward in an FFNN, errors are propagated backward while adjusting the weights and biases according to the error. The back propagation algorithm uses the concept of supervised learning with error correction and minimization to train the network, which means that in a given sample of inputs and target or desired output to the network. Error is calculated between the target and actual output (S. N. Kavitha, C. T. S, and P. Jannu, 2018). The idea of a back propagation algorithm is to obtain the target as output when a respective input is given. The error depends on the weights and biases in the neurons. Then the weights and biases are adjusted in order to minimize the output error.

2.4.2 Related works

Prediction of Malaria using artificial neural network with MPL (multi-layer perceptron) proposed by (Rahila parvem,Akhtar Hussain,et.al. 2017). is used along with back propagation, back propagation with momentum resilient propagation rule for the prediction of malaria. Among all three learning rules back propagation gives the more efficient results that are approximately 85% accuracy. The history and symptom of a patient are considered as an input. The system analysis that data and predict the result for the victim as positive or negative for malaria. These is important and useful where there is no laboratory facility or where there is no doctor because the system generates the type of the disease after accepting the symptom as input. In this proposed system there is not considered a cause of malaria, factors for malaria by relating mosquito Vector with climatic factor for malaria abundance including environmental and clinical factor for creating more awareness to the people.

Malaria incident Forecasting investigated by (Anditya Arifianto.et.al. 2014) that apply Polynomial Neural Network is used to reduce the learning time and computation while maintain the accuracy in predicting malaria incidence by relating it to weather pattern. Based on the experiment it was proven that the modified GMDH PNN was able to reduce the learning time by 72% and improve the accuracy in to 88.02% compared to the original GMDH PNN. In this method malaria incident forecasting is the problem of time series prediction and a time series prediction involves the determination of an appropriate model which can encapsulate the dynamic of the system described by the sample data. Inductive GMDH algorithm gives the possibility to find automatically interrelation in data to select an optimal structure of a model to increase the accuracy of the existed algorithm. This model is not really important and applicable in real-world practice because the goal of this model is by taking some sample data, use some learning and training time and gate better accuracy in addition to reduce learning time.

The prediction Methods of malaria are proposed by (Thakur setosh,Dharauth Ramesh. 2018). In those Method data collected from PHC of the department of Vector-Borne diseases (DVBD) of specific geographical location in India. The prediction model is developed using the feed-forward neural network (FFNN) and compared to the observed value. In those proposed algorithms, there is maximum error accord in the expected

output in the oldest algorithm.

Prediction of mosquito abundance using ANN implemented by (Keun young lee,Namil chung et al. 2015). These proposed method are used ANN and MLR methods that are used for prediction by counting the number of mosquito in two years of data by different season using digital mosquito Monitoring system in Urban areas. The input variable are metrological variable in Korea seole, those model used for performance evaluation of compared ANN and MLR, then according to the predicted result ANN result is better performance than MLR, the model is not give accurate and best fit to practical realization because the sample are only two years and the input variables are only metrological variable for counted the Mosquito abundance relative to weather condition in specific location of seol. Additionally the proposed system does not explore malaria abundance and the study is only mosquito does not consider the disease. This implementation does not identify the malaria disease.

Generic Weather Driven Model to predict Mosquito population Dynamics proposed by (P.Ezanno,M.Aubrykientz et al. 2015)**,** those work presents an accurate understanding and prediction of mosquito population dynamics are needed to identify areas where there is a high risk of mosquito-borne disease spread and persistence. Simulation tools are relevant for supporting decision-makers in the surveillance of vector population.

As models of vector population dynamics provide predictions of the greatest risk periods for vector abundance, which can be particularly helpful in areas with a highly variable environment. Those present a generic weather-driven model of mosquito population dynamics for prediction, which was applied to one species of each of the genera *Anopheles*, *Culex*, and *Aedes* are located in the same area and thus affected by similar weather conditions.The data were previously obtained using CDClight traps baited with carbon dioxide dry ice in 2005 at two sites in a favorable temperate wetland of southern France. The observed and predicted periods of maximal abundance for *An. Hyrcanus* and *Cx. pipiens* tallied very well, those model does not include a spatial component, in a favorable environment with a mosquito. There is not clinical condition are used for sample the data for prediction. But the model was mechanistic and driven by meteorological factors, with transitions depending on temperature and rainfall occurrence and intensity. It also represented the seasonal dynamics of a mosquito population over several years, accounting for diapauses occurring in relation to day- length. The model for *An. hyrcanus* predicted this species to be particularly abundant over quite a short period of time in late summer, whereas the model for *Cx. pipiens* predicted this species to be less abundant but over a longer period of time from early summer up to the end of the favorable

period. This proposed system gives spatial attention to mosquitoes using their own counting mechanism by their own model and does not create awareness about malaria disease and their transmission does not give good prediction using ANN.

The Malaria Outbreak Prediction Model Using Machine Learning are proposed by (Vijeta Sharma1, Ajai Kumar2 et. al. **(**2015), by this study only classify has used to choose Support Vector Machine (SVM) and Artificial Neural Network (ANN) classification algorithms. Performance of chosen classifier on dataset shows as WEKA's statistical output in classifier output panel, by analyzing different performance parameters. Classifier can be the best for a given dataset, that is SVM. in this proposed work the main objective are compare SVM with Neural network model in addition to another limitation.

Detection of mosquito with NN model was investigated by (Ivan Kiskin1;2, Bernardo P_erez Orozco1;2, et.al. 2017), the methods are present a novel wavelet-transform-based convolutional neural network architecture for the detection of mosquitoes' in a noisy audio recording, those explain the wavelet transform in the context of the algorithm, there after describing the neural network configurations and a range of traditional classifiers against which we assess performance. The key steps of feature extraction and classification pipeline are given in the Algorithm. Those proposed emphasize that the ultimate goal is deployment in fieldwork on smartphones or embedded devices. The device will be in constant listening mode, and mainly consume power during the data write mode that is initiated by signal detections, a high true negative rate (TNR) is very desirable for this application, as preventing false positive detections leads to critical conservation of battery power, taking this into account, these highlight four key results. This work concludes the addition of a convolutional layer results in performance gains over non-convolutional neural networks with both Fourier and Wavelet representations.

From previous related works, they mostly study the mosquito and malaria in different place of the world by different researchers and different methodology to trainee and predict in machine learning. This thesis addresses and develop the LSTM model, which is one of the machine learning recent algorithm used for trainee, test and predict the amount of malaria in amhara region using python simulation tools. In addition to amhara region in the geographical Region of South Gondar and West Go jam zone are also used in these study, those are the geographic part of Amhara region that included and selected for malaria abundance prediction using the developed model and tools.

CHAPTER THREE

3. METHODOLOGY

3.1 Introduction

This chapter presents the design and implementation of the proposed model. This chapter presents the design and implementation of the proposed model. The overall methodology used in the research is presented using a flowchart as shown in Figure 3.1

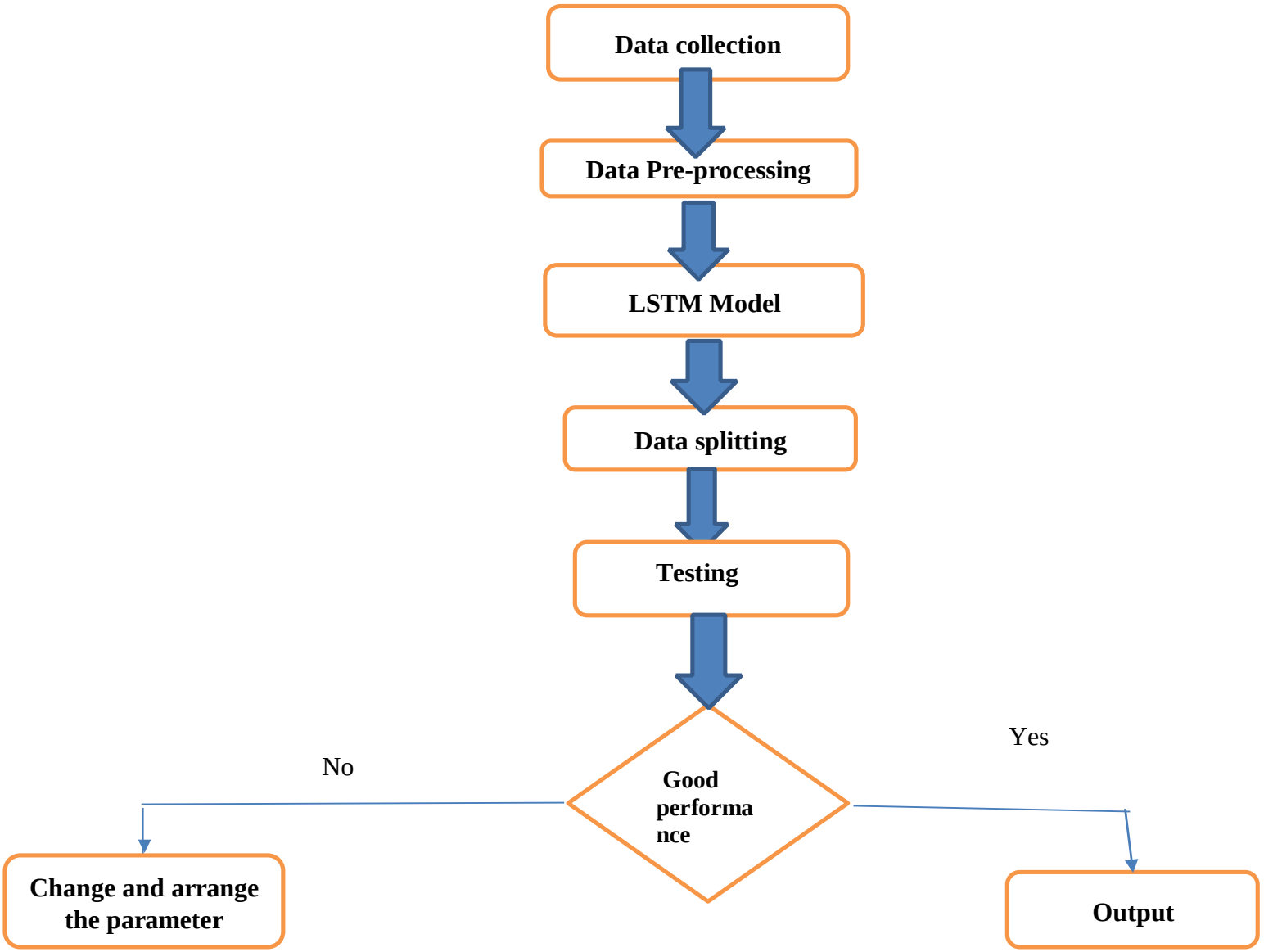


Figure 3.1: System flowchart

3.2. Study area and Data Collection

3.2.1 Study Area

This research is conducted in Ethiopia Amhara regional state. Amhara region is one of the nine geographic regions in Ethiopia and is located in the north-west of the country, covering approximately 150,000 km, and also the second most populated state that comprises a number of districts of woredas and zones. The highlands receive 80% of the total rainfall of Ethiopia and is the most fertile and hospitable region of Ethiopia(W. B. Abegaz and A. Mekoya,2020). The Amhara region is the location of Lake Tana, the source of the Blue Nile, at Bahir Dar. The flow of the Blue Nile reaches maximum volume in the rainy season (from June to September) when it supplies about two-thirds of the water of the Nile proper.

The Following shows study area location

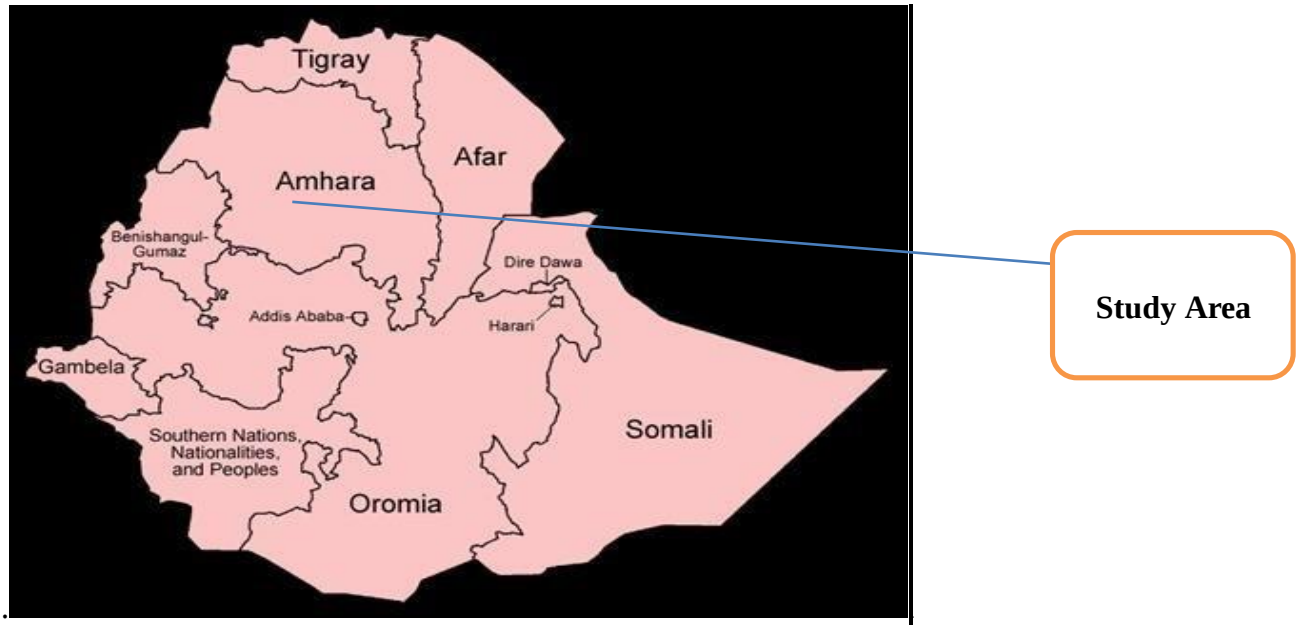


Figure 3.2: Study area location

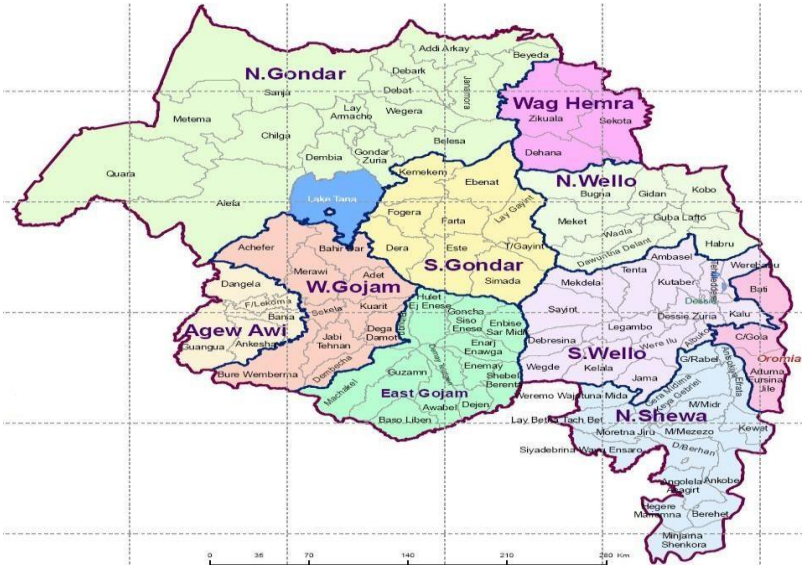


Figure 3.3: Specific study location

3.2.2 Data Collection

This research metrological data are taken from Bahir Dar meteorological branch in Bahir Dar that consists of three parameter like relative humidity, temperature, rainfall, and the clinical data taken from the Amhara public health institute. Those metrological data are minimum temperature, maximum temperature, rainfall and humidity, Additionally the malarial cases are used for this work:

- **Relative humidity (RH):** is the ratio of water vapor contained in the air compared to the maximum amount of moisture that the air can hold at that specific temperature and pressure. Relative humidity depends on temperature and the pressure of the system of interest. The same amount of water vapor results in higher relative humidity in cool air than warm air.
- **Temperature:** This is the amount of the air's hotness or coldness and measured in degrees on the Fahrenheit, Celsius, and Kelvin scales.
- **Minimum Temperature:** The lowest temperature recorded daily, monthly, seasonally, or annually, or the lowest temperature of the entire record.
- **Maximum Temperature:** This is the highest temperature recorded during a specified period of time. Common time periods include 6, 12, and 24 hours. The most common reference is to the daily maximum temperature.

- **Malaria cases:** This is the number of malaria patient that causes by anopheles mosquito in a given period of time.
- **Rainfall:** This is a type of precipitation that occurs when water vapor in the atmosphere condenses into droplets that can no longer be suspended in the air. The occurrence of rainfall is dependent upon several factors. Things such as prevailing wind directions, ground elevation, and location with respect to mountain ranges.
- The Malarial cases are taken as output and the rest of the parameters are taken as the input of the model. In the station, the metrological parameter was recorded about 12 years (20062017). For this research, these data are used for study and forecasting purposes. The data for the parameters are shown below.

Table 3.1: Amhara region monthly malarial cases (2006-2017)

year/month	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July	Aug
2006	**245771**	**60188**	**100318**	**20062**	**60188**	**12037**	**102518**	**80251**	**12037**	**60188**	**40123**	**140440**
2007	**550172**	**450172**	**411129**	**451422**	**20396**	**25597**	**17590**	**68563**	**133155**	**200459**	**58986**	**45970**
2008	**207456**	**208511**	**120619**	**108527**	**88632**	**60781**	**30321**	**22625**	**202401**	**305896**	**100965**	**978596**
2009	**120241**	**90896**	**70265**	**66983**	**20478**	**17896**	**36987**	**60983**	**90983**	**120938**	**42890**	**55789**
2010	**120765**	**94765**	**83770**	**61768**	**52260**	**37658**	**24770**	**24765**	**70768**	**99765**	**30765**	**50982**
2011	**40390**	**36688**	**35080**	**34393**	**14000**	**13900**	**10690**	**10090**	**39385**	**41395**	**30390**	**34765**
2012	**41650**	**32053**	**28340**	**24348**	**18748**	**2008**	**9356**	**9940**	**22350**	**41348**	**29348**	**30765**
2013	**36516**	**34516**	**32016**	**30016**	**11016**	**10016**	**10522**	**20510**	**32016**	**36016**	**26522**	**25765**
2014	**8466**	**7436**	**6836**	**3900**	**4006**	**3036**	**2866**	**3001**	**7131**	**9036**	**8436**	**20765**
2015	**36083**	**22006**	**21583**	**19800**	**11039**	**10247**	**13040**	**12046**	**30038**	**24048**	**22043**	**26765**
2016	**20152**	**25698**	**45726**	**64077**	**68958**	**2356**	**5566**	**7892**	**35955**	**89751**	**57536**	**8954**
2017	**63165**	**9865**	**8869**	**65312**	**9875**	**48991**	**6526**	**9684**	**86985**	**79856**	**3652**	**98456**

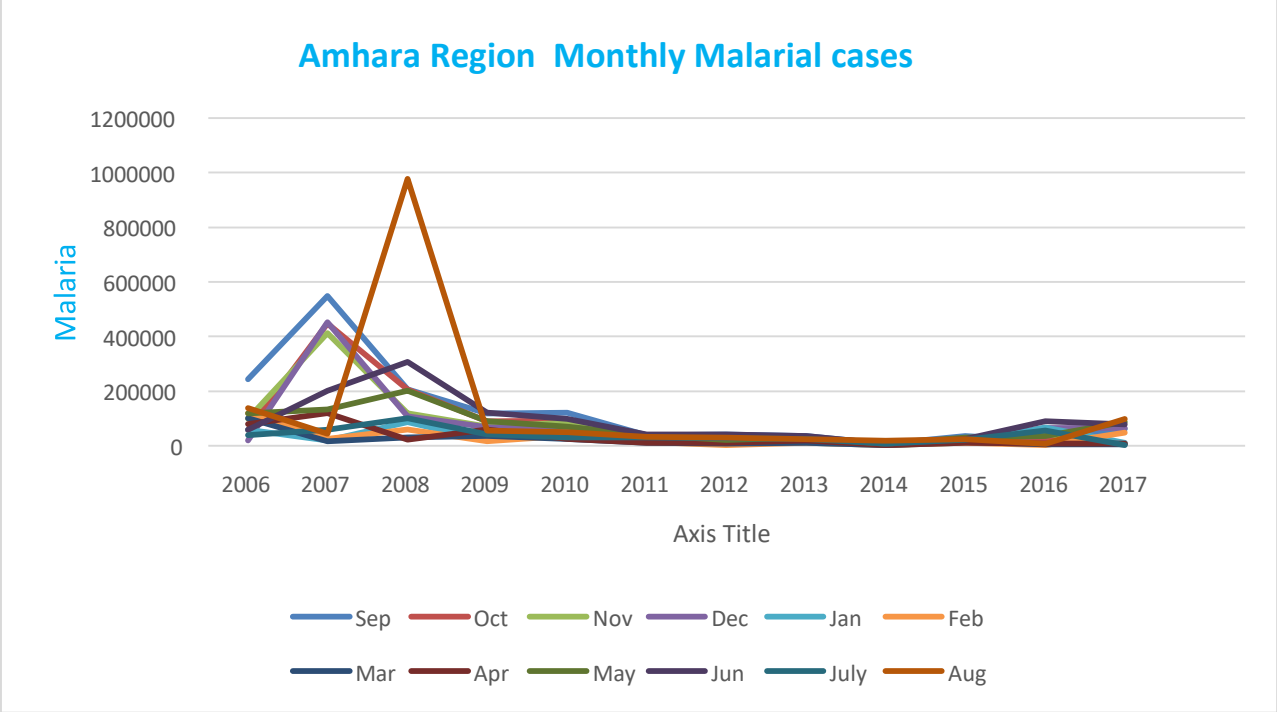


Figure 3.4: Amhara region monthly malaria cases in chart (2006-2017)

Table 3.2: Amhara region monthly maximum temperature (2006-2017)

Monthly Maximum
Element: temperature

Year/month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	31.5	32.5	32.5	32.0	30.0	26.0	25.0	24.5	29.3	27.0	25.5	20.0
2007	28.5	36.6	31.5	37.5	32.0	25.5	24.5	27.0	25.0	27.0	29.0	29.5
2008	28.0	30.4	30.5	35.6	31.0	23.0	25.0	29.0	23.3	28.1	25.2	28.3
2009	29.3	28.9	30.0	29.0	30.0	24.0	26.0	28.7	24.3	28.3	26.0	28.0
2010	30.0	31.0	30.0	33.0	28.6	24.0	22.0	28.0	29.0	26.0	24.0	27.0
2011	30.0	32.0	31.0	29.5	32.0	26.0	25.0	27.0	30.0	35.1	36.7	36.3
2012	31.0	30.3	33.0	31.3	39.6	34.1	30.4	30.6	31.7	36.3	36.9	36.6
2013	37.5	39.9	40.8	41.3	39.2	34.7	31.2	29.4	32.3	34.1	37.1	36.0
2014	37.2	37.5	39.8	39.7	38.4	28.5	25.0	23.5	26.6	27.0	27.5	36.8
2015	31.52	40.2	41.1	40.3	38.0	41.1	44.0	31.5	32.2	34.7	36.7	34.6
2016	30.3	38.4	41.8	41.1	37.3	35.4	30.7	30.5	33.8	34.3	37.1	37.9
2017	38.8	37.1	40.3	40.5	37.4	35.1	31.6	30.0	31.8	34.4	36.4	37.3

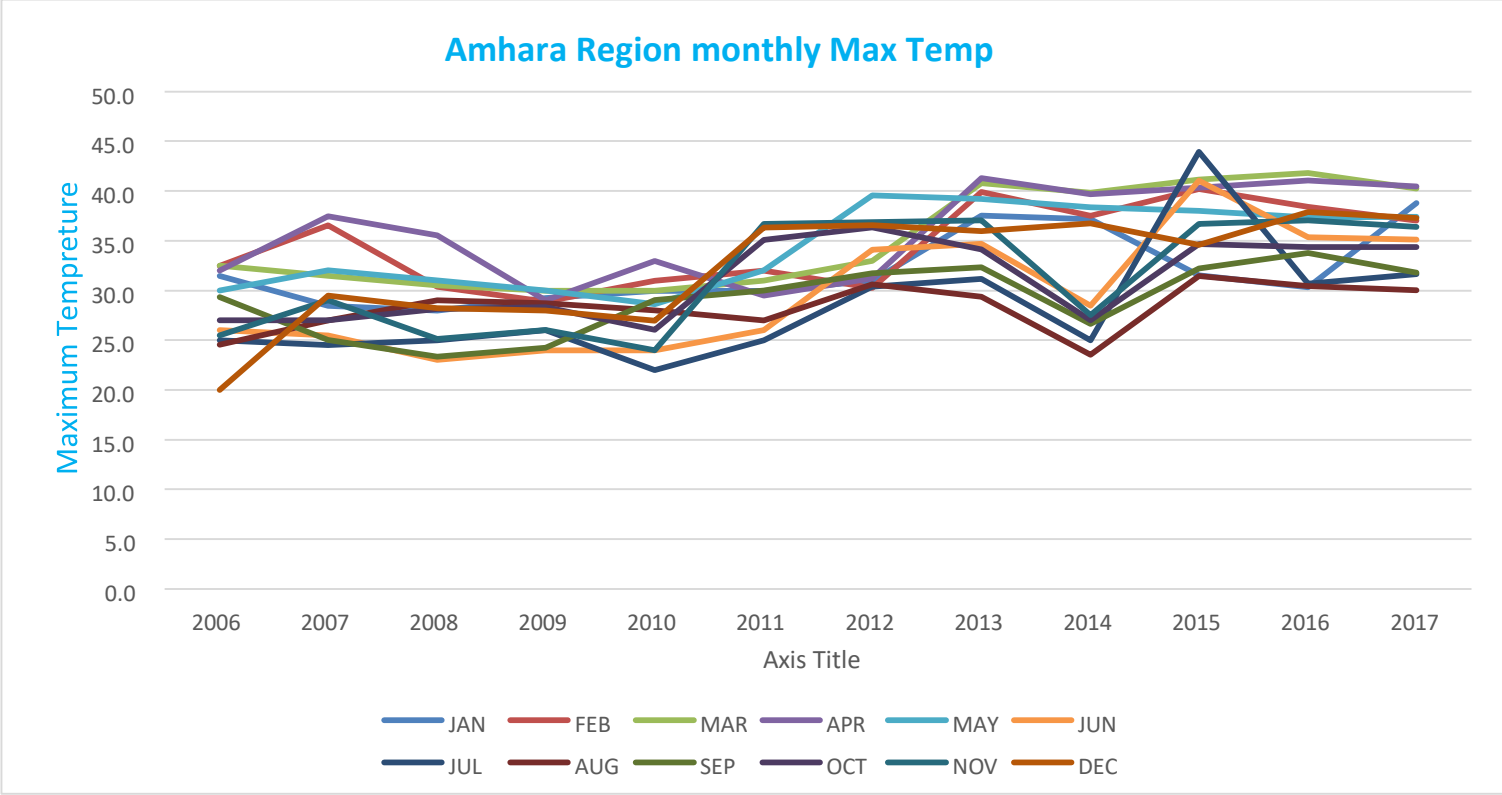


Figure 3.5: Amhara region monthly maximum temperature in chart (2006-2017)

Table 3.3: Amhara region monthly relative humidity (2006-2017)

	Amhara Region monthly RH											
MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	43	39	36	38	62	70	81	81	69	55	46	45
2007	53	57	47	48	52	67	79	79	66	64	58	49
2008	47	51	43	50	59	79	70	78	79	73	65	56
2009	34	44	43	40	43	72	56	67	181	71	70	61
2010	50	43	38	43	59	68	80	82	77	66	58	54
2011	55	34	45	40	57	70	76	81	77	63	60	51
2012	43	33	37	30	46	61	76	78	75	62	63	55
2013	46	37	36	34	53	65	77	82	74	70	58	47
2014	45	41	47	51	66	67	76	77	74	67	55	51
2015	49	39	38	35	59	68	77	75	74	65	61	65
2016	54	44	41	39	63	69	80	77	75	66	49	45
2017	37	53	42	47	64	67	76	78	73	69	58	76

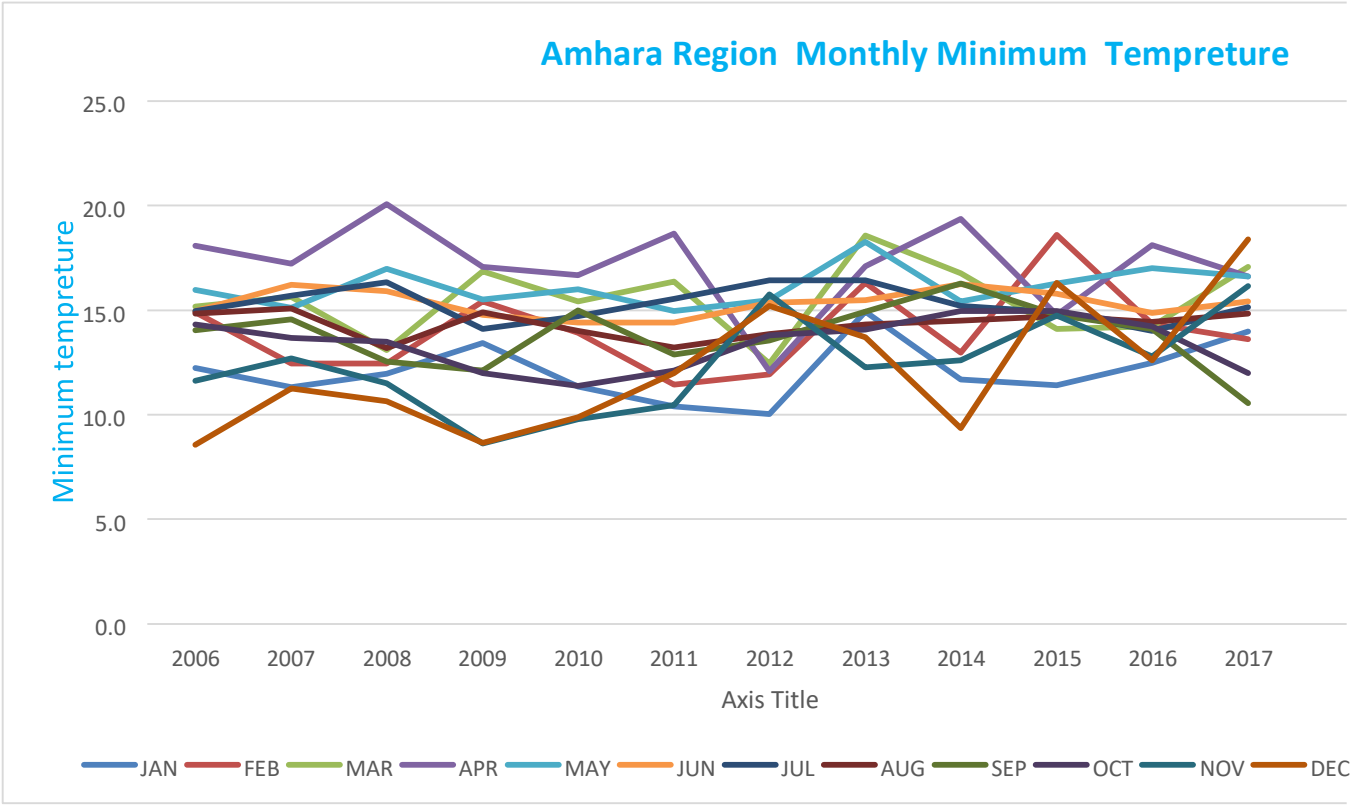


Figure 3.6: Amhara region monthly Minimum Tempreture in chart (2006-2017)

Table 3.4: Amhara region monthly Relative Rainfall (2006-2017)

Amhara Region Monthly Rainfall

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	0.0	0.0	10.8	27.8	152.6	98.7	291.5	305.3	192.5	87.4	29.9	35.0
2007	4.8	0.0	15.8	45.1	101.6	162.2	340.7	355.1	126.6	96.2	10.6	0.0
2008	1.7	0.0	0.0	63.5	104.7	228.5	365.6	301.5	106.6	15.7	17.9	15.1
2009	1.8	12.5	2.7	12.6	8.0	195.2	293.0	300.4	58.2	46.1	3.2	14.8
2010	19.7	0.0	23.8	45.4	145.9	105.4	266.6	325.0	71.7	37.4	14.6	0.0
2011	2.5	0.0	19.9	23.8	104.7	172.0	230.9	268.0	103.5	52.3	46.6	0.0
2012	9.8	0.0	8.9	0.0	31.2	117.3	381.5	354.4	135.2	43.9	51.1	11.1
2013	0.0	0.0	1.4	1.6	88.0	148.6	594.0	350.3	137.9	169.1	16.0	0.0
2014	0.0	0.0	65.9	66.6	163.7	178.4	378.4	480.8	260.0	117.4	0.0	0.4
2015	0.0	0.8	0.4	0.0	136.8	89.3	302.2	248.9	223.9	116.7	12.2	31.0
2016	0.0	0.0	23.8	8.5	171.2	248.8	409.6	259.9	274.4	104.8	0.5	0.0
2017	0.0	25.9	2.6	41.9	136.8	50.0	528.7	485.7	189.2	128.9	9.6	0.0

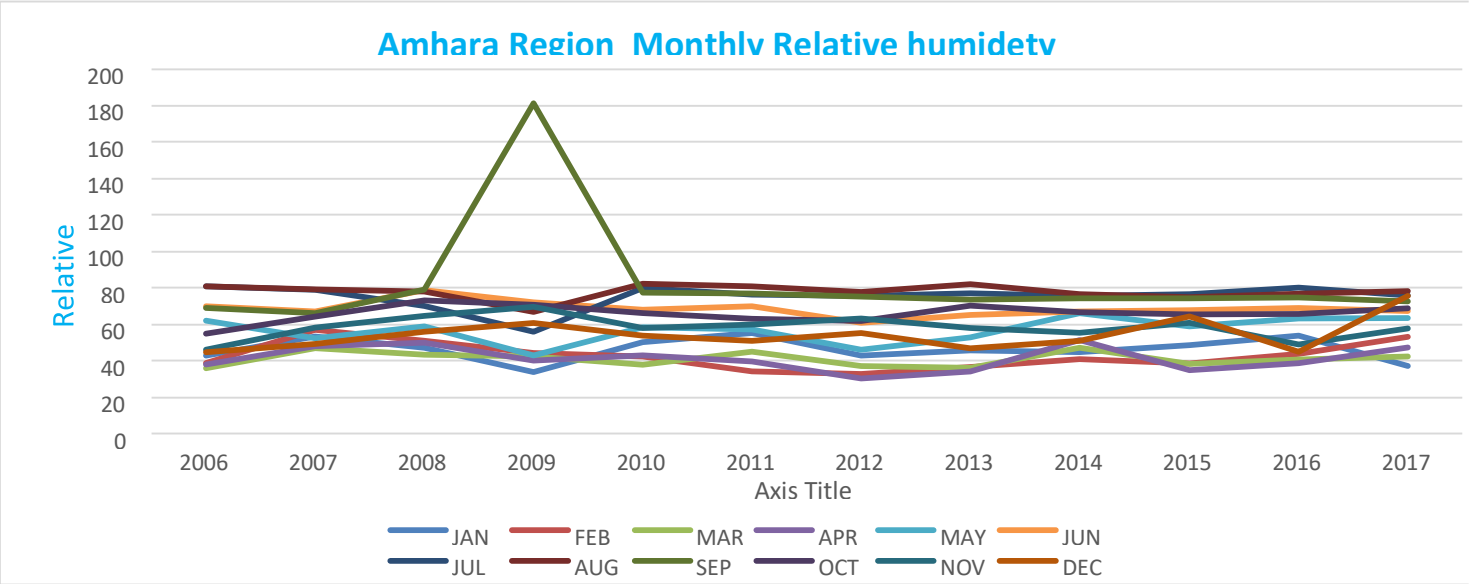


Figure 3.7: Amhara region monthly Relative Humidity in chart (2006-2017)

Table 3.5: Amhara Region Monthly Minimum Tempreture (2006-2017)

				Amhara Region							**LAT.**	**11.590**
				Element : Monthly Minimum temprature							**ELEV.**	**180**
MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	12.2	14.9	15.2	18.1	16.0	15.0	15.0	14.8	14.1	14.3	11.6	8.6
2007	11.3	12.5	15.6	17.2	15.1	16.2	15.7	15.1	14.6	13.7	12.7	11.3
2008	12.0	12.5	13.1	20.1	17.0	15.9	16.3	13.2	12.5	13.5	11.5	10.6
2009	13.4	15.4	16.9	17.1	15.5	14.8	14.1	14.9	12.1	12.0	8.6	8.7
2010	11.4	13.9	15.4	16.7	16.0	14.4	14.7	14.0	15.0	11.4	9.8	9.9
2011	10.4	11.4	16.4	18.7	15.0	14.4	15.6	13.2	12.9	12.1	10.5	12.0
2012	10.0	11.9	12.5	12.1	15.5	15.4	16.4	13.9	13.6	13.8	15.8	15.2
2013	14.9	16.3	18.6	17.1	18.3	15.5	16.4	14.3	14.9	14.1	12.3	13.7
2014	11.7	13.0	16.8	19.4	15.4	16.3	15.2	14.5	16.3	15.0	12.6	9.4
2015	11.4	18.6	14.1	14.8	16.3	15.8	14.9	14.7	14.8	15.0	14.7	16.3
2016	12.5	14.3	14.2	18.1	17.0	14.9	14.0	14.4	14.1	14.2	12.8	12.6
2017	14.0	13.6	17.1	16.6	16.6	15.4	15.1	14.9	10.6	12.0	16.1	18.4

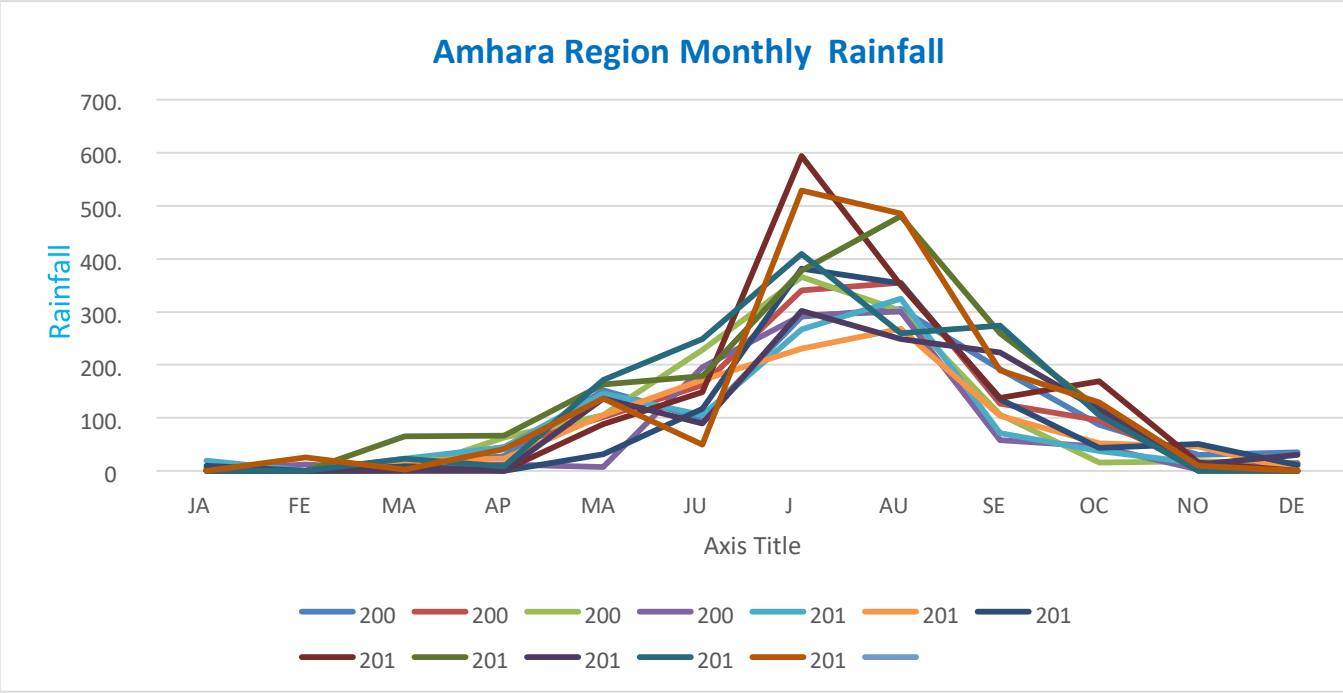


Figure 3.8: Amhara region monthly Relative Rainfall in chart (2006-2017)

3.3 Data Preprocessing

Data preprocessing is an important step in the data mining and machine learning process. The phrase "garbage in, garbage out" is particularly applicable to both fields. As data gathering techniques are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. The process of data preparation for a machine-learning algorithm includes three steps:

Select data: Consider what data is available, what data is missing, and what data can be removed

Preprocess Data: Organize the selected data by formatting, cleaning, and sampling from it

Transform Data: Transform preprocessed data ready for feed machine learning model by engineering features using scaling and attribute aggregation

3.3.1 Missing Values handling

The real-world data often has a lot of missing values. The cause of missing values can be data corruption or failure to record data. The handling of missing data is very important during the preprocessing of the dataset,

as many machine-learning algorithms do not support missing values. The method to handle this problem is to get rid of the observations that have missing data. There are different methods to handle missing values:
Do Nothing: That is an easy one. Just let the algorithm handle the missing data. Some algorithms can factor in the missing values and learn the best imputation values for the missing data based on the training loss reduction. Some others have the option to just ignore them so that the missing data are handled and clean it before feeding it to the algorithm
Imputation Using(Mean/Median) values: These techniques perform by calculating the mean/median of the non-missing values in a column and then replacing the missing values within each column separately and independently from the others But these techniques is used for numeric data only

Imputation using (Most Frequent) or (Zero/constant) Values: This works by imputing all missing values with **zero**, the **mean** or **median** for quantitative variables, or the most common value for categorical variables.

Imputation with a model

These are worked by replacing missing values with predicted values from a model based on the other observed predictors. The *k* nearest neighbor algorithm is often used to impute a missing value based on how closely it resembles the points in the training set.

Model-based imputation with uncertainty: works by replacing missing values with predicted values plus randomness from a model based on the other observed predictors.

Model-based progressive imputation: uses previously imputed missing values to predict other missing values.

In these weather datasets, if any missing data occur we can handle using median/mean values because the data are numeric attribute and missing values of time series data have a strong relation to its previous and values in the series.

3.3.2 Normalization

Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling. Data transformation should be done to improve the speed, accuracy, efficiency, and performance of the model. The performance of a neural network is improved through data transformation (normalization). According to the literature, there are three existing data transformation methods.

Linear transformation, Statistical standardization, and Mathematical function. The most commonly used one is scaling are linear transformation methods. In this work scaling based data normalization is used to transform the values of all attributes in the weather dataset to [0,1] range. The normalization is done as follows

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where X'=is the normalized value, X is the original values, X min is the minimum value and X max is the maximum value of the original data.

Table 3.6: Amhara Region Monthly Normalized Minimum temperature (2006-2017)

year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	0.44685	0.401027	0.442836	0.751646	0.30826	0.303797	0.392053	0.864407	0.611208	0.822144	0.400177	0
2007	0.26378	0.118485	0.518581	0.644612	0.045675	0.974684	0.690066	1	0.701693	0.638747	0.542275	0.274574
2008	0.392388	0.119043	0.104613	1	0.607386	0.81736	0.952318	0	0.346176	0.590579	0.382913	0.211664
2009	0.690945	0.460684	0.717938	0.625713	0.162293	0.197107	0.038411	0.90339	0.272037	0.171391	0	0.009666
2010	0.269685	0.287513	0.482812	0.573952	0.311953	0	0.298013	0.423729	0.778751	0	0.155378	0.135321
2011	0.078084	0	0.638511	0.825233	0	0.005861	0.635762	0.013559	0.405721	0.207331	0.245241	0.35059
2012	0	0.056643	0	0	0.156715	0.511754	0.996026	0.348418	0.525978	0.672255	0.947322	0.675644
2013	1	0.564256	1	0.627595	1	0.587503	1	0.59661	0.763962	0.751239	0.485613	0.523309
2014	0.335302	0.178465	0.701632	0.912616	0.141885	1	0.496689	0.690482	1	0.996266	0.53112	0.081021
2015	0.284121	0.83107	0.268727	0.33856	0.398445	0.754069	0.364238	0.805085	0.738179	1	0.813192	0.786697
2016	0.497878	0.336242	0.289767	0.755744	0.616132	0.24774	0	0.658418	0.618797	0.79085	0.5529	0.41055
2017	0.803675	0.251885	0.751603	0.569771	0.499514	0.549729	0.468874	0.871186	0	0.173165	1	1

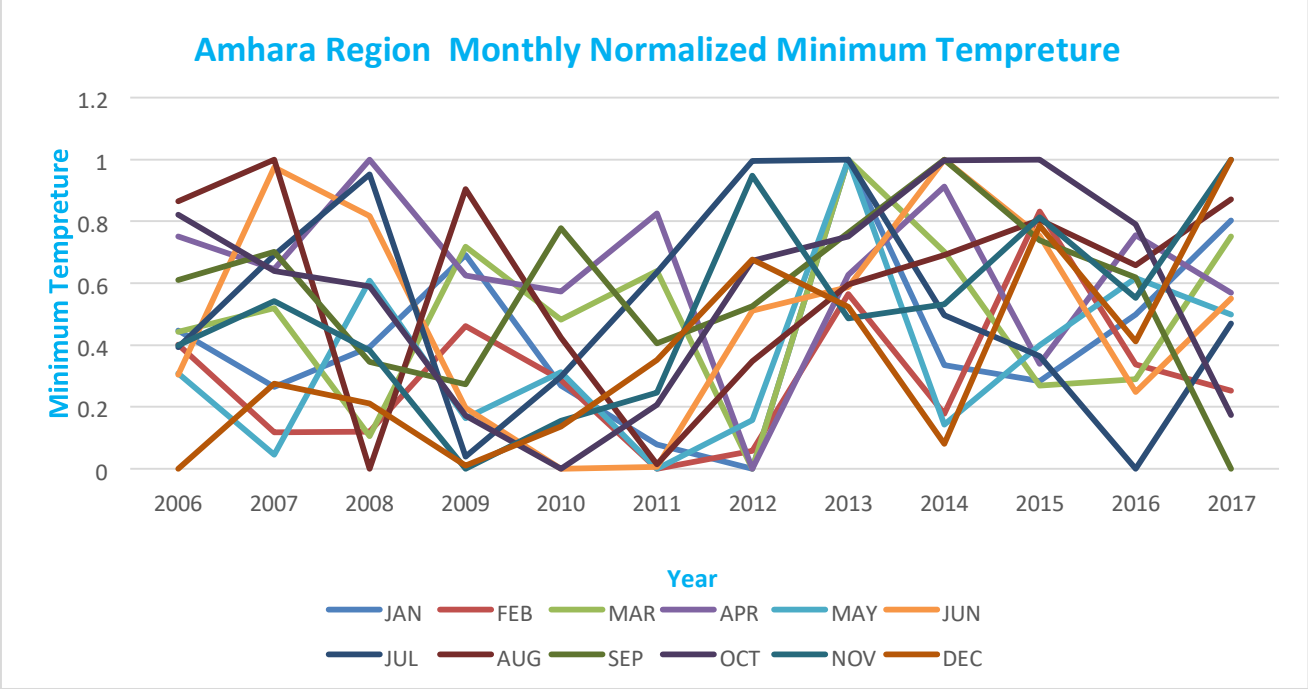


Figure 3.9: Amhara region monthly-normalized minimum temperature in chart (20062017)

Table 3.7: Amhara region monthly normalized relative rainfall (2006-2017)

Year/month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	0	0	0.1639	0.417	0.886	0.245	0.167	0.24	0.6212	0.467	0.59	1
2007	0.2437	0	0.2398	0.677	0.574	0.564	0.302	0.45	0.3164	0.525	0.21	0
2008	0.0863	0	0	0.953	0.593	0.898	0.371	0.22	0.2239	0	0.35	0.4314
2009	0.0914	0.5	0.041	0.189	0	0.73	0.171	0.22	0	0.198	0.06	0.4229
2010	1	0	0.3612	0.682	0.845	0.279	0.098	0.32	0.0624	0.141	0.29	0
2011	0.1269	0	0.302	0.357	0.593	0.614	0	0.08	0.2095	0.239	0.91	0
2012	0.4975	0	0.1351	0	0.142	0.339	0.415	0.45	0.3562	0.184	1	0.3171
2013	0	0	0.0212	0.024	0.49	0.496	1	0.43	0.3686	1	0.31	0
2014	0	0	1	1	0.954	0.646	0.406	0.98	0.9334	0.663	0	0.0114
2015	0	0	0.0061	0	0.789	0.198	0.196	0	0.7664	0.658	0.24	0.8857
2016	0	0	0.3612	0.128	1	1	0.492	0.05	1	0.581	0.01	0
2017	0	1	0.0395	0.629	0.789	0	0.82	1	0.6059	0.738	0.19	0

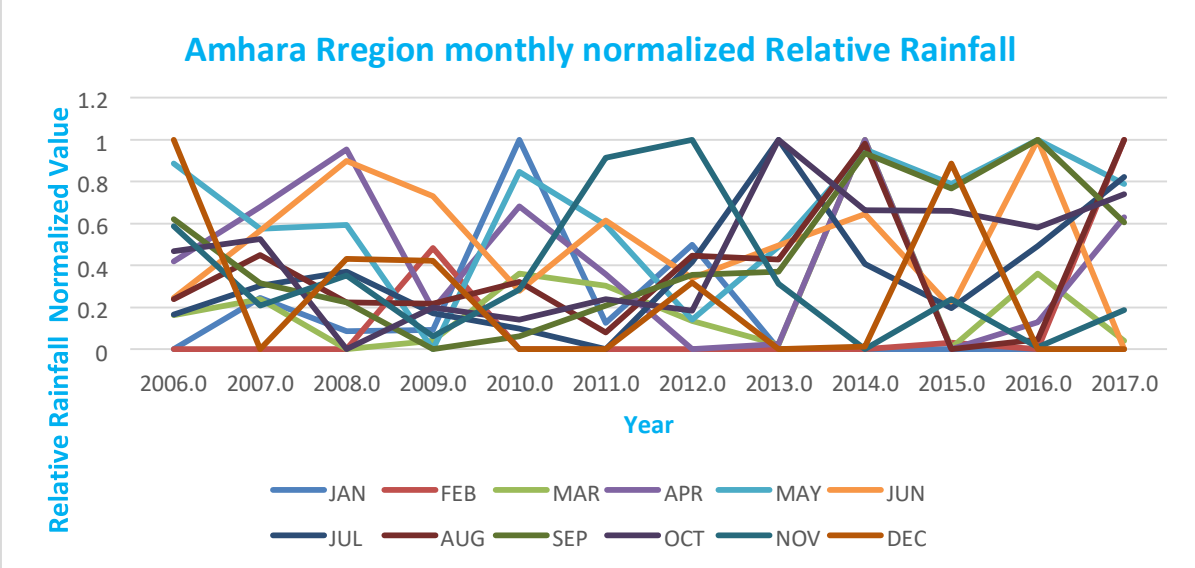


Figure 3.10: Amhara region monthly normalized relative rainfall in chart (2006-2017)

Table 3.8: Amhara region monthly normalized relative humidity (2006-2017)

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	0.419019	0.249258	0	0.369255	0.817294	0.505618	1	0.910603	0.02521	0	0	00
2007	0.907875	1	0.974359	0.846276	0.405858	0.346442	0.928756	0.798337	0	0.516934	0.511173	0.14418
2008	0.624071	0.754651	0.655271	0.939778	0.690377	1	0.567358	0.721414	0.11417	1	0.789106	0.364134
2009	0	0.470326	0.578348	0.472266	0	0.629213	0	0	1	0.86631	1	0.522206
2010	0.766716	0.39911	0.176638	0.606973	0.669456	0.393258	0.959845	1	0.097653	0.618538	0.502793	0.293622
2011	1	0.053412	0.803419	0.448494	0.610879	0.503745	0.816062	0.910603	0.094755	0.442068	0.586592	0.195748
2012	0.424963	0	0.101403	0	0.136681	0	0.78886	0.715177	0.078818	0.377897	0.72905	0.339928
2013	0.552749	0.15727	0.017094	0.179081	0.432357	0.234082	0.843264	0.987526	0.064619	0.843137	0.502793	0.066302
2014	0.514116	0.338279	1	1	1	0.338951	0.786269	0.638254	0.071573	0.642424	0.394972	0.195748
2015	0.688815	0.239614	0.204416	0.221257	0.68424	0.391386	0.830311	0.527651	0.071458	0.577184	0.628492	0.636287
2016	0.928975	0.4564	0.420085	0.402932	0.87643	0.449438	0.969326	0.652807	0.076384	0.589162	0.12067	0.002526
2017	0.160475	0.84451	0.567949	0.806656	0.883961	0.355805	0.80285	0.747609	0.056447	0.762923	0.493017	1

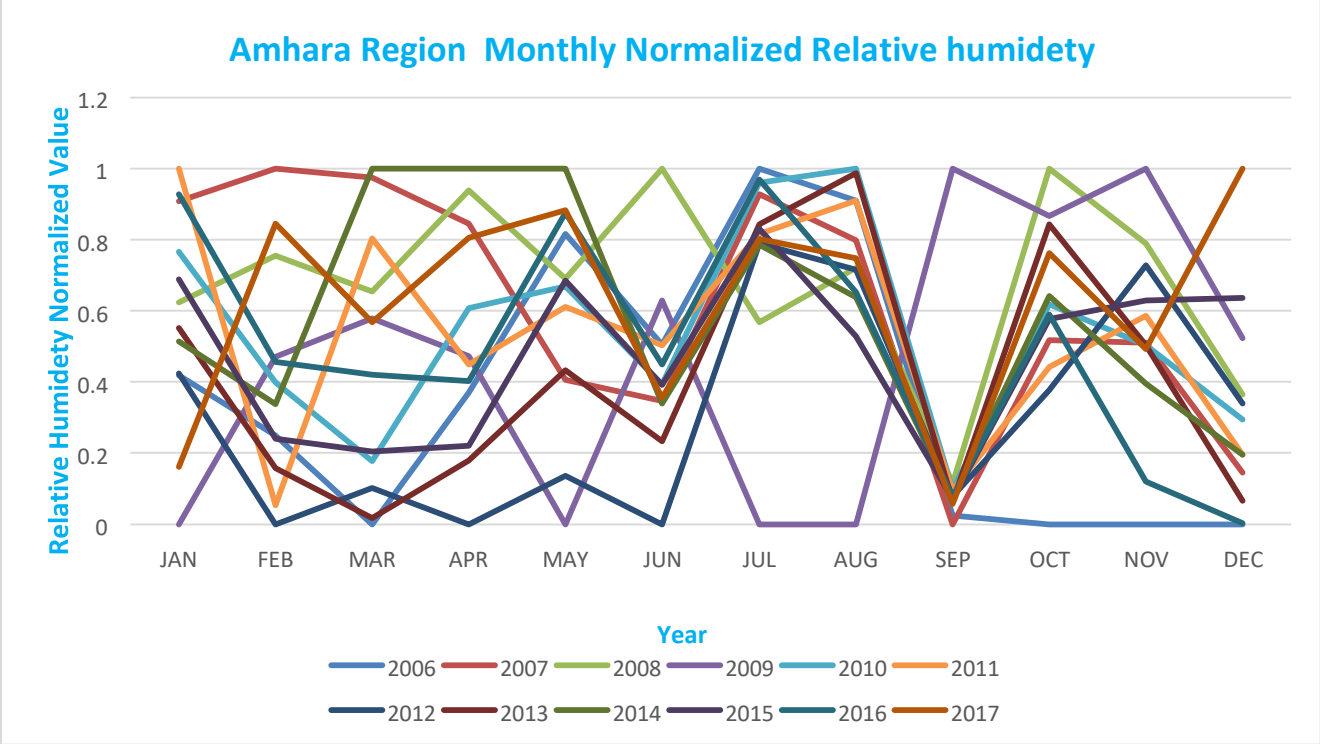


Figure 3.11: Amhara region monthly normalized relative humidity in chart (2006-2017)

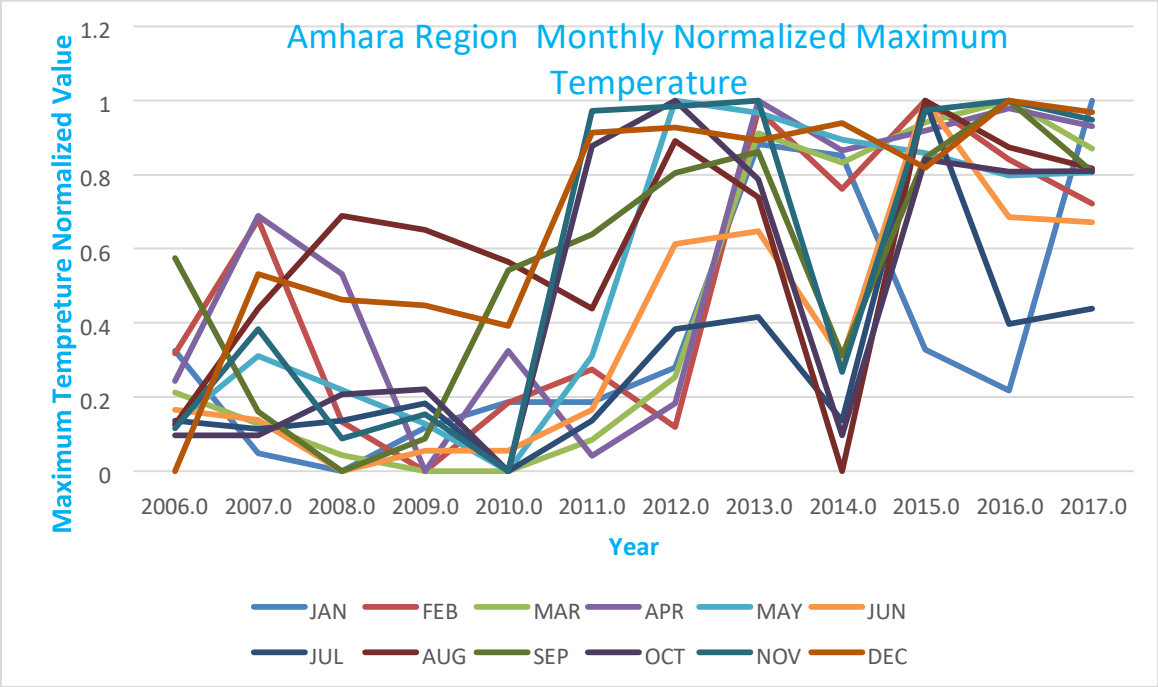


Figure 3.12: Amhara region monthly-normalized maximum temperature in chart (2006-2017)

Table 3.9: Amhara region monthly-normalized malaria cases (2006-2017

year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	0.663886	1	1	0.659293	0.579946	0.17231	0.37478	0.135603	0.432907	0.11915	0.231223	0.036114
2007	0.193676	0.199282	0.147754	1	0.645383	0.644826	0.568619	0.038175	0.988215	1	1	1
2008	1	0.496519	0.275509	0.167482	1	1	1	1	0.36301	0.454165	0.281437	0.233792
2009	0.194645	0.134223	0.342402	0.494849	0.429416	0.376952	0.403214	0.048301	0.203907	0.18851	0.156889	0.140961
2010	0.570203	0.301174	0.219805	0.185746	0.325892	0.305629	0.278616	0.043344	0.204863	0.197248	0.190293	0.129308
2011	0.118096	0.100465	0.078513	0.060501	0.165176	0.109004	0.274763	0.026619	0.058238	0.066071	0.06986	0.068137
2012	0.174202	0	0.065127	0.059221	0.077938	0.108846	0.264055	0.022494	0.060536	0.055602	0.053189	0.045692
2013	0.082835	0.067652	0.076827	0.149431	0.127439	0.090885	0.235015	0.017337	0.051171	0.061165	0.062282	0.058357
2014	0	0.008685	0	0	0	0	0.049161	0.012181	0	0	0	0
2015	0.083107	0.069604	0.102095	0.077195	0.117309	0.050569	0.188988	0.018369	0.050381	0.032909	0.036476	0.035529
2016	0.767518	0.00294	0.027094	0.041742	0.147611	0.271896	0.553718	0	0.021318	0.041248	0.096193	0.134467
2017	0.069352	0.396916	0.036728	0.057036	0.408941	0.238564	0	0.092304	0.099785	0.005486	0.005029	0.137227

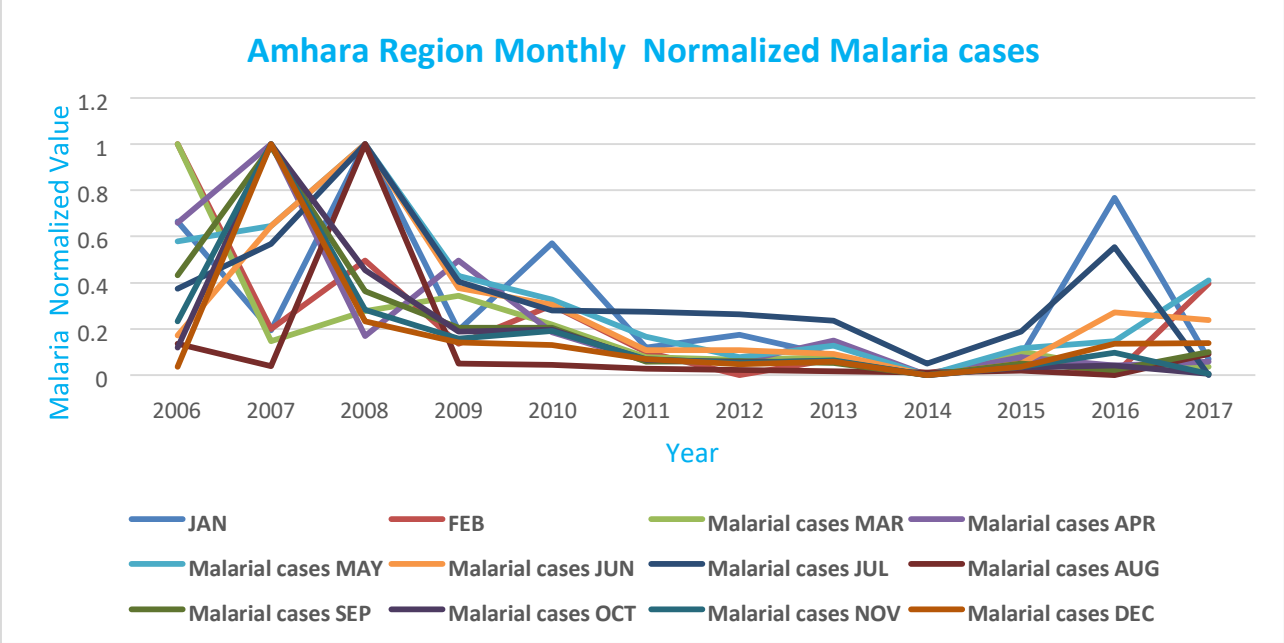


Figure 3.13: Amhara region monthly normalized malaria cases in chart (2006-2017)

So all these the above figure are normalized and un normalized values of the weather and malarial attributes that include in these studies. After Normalization all, these attributes are merged for a common goal to one chart as shown below

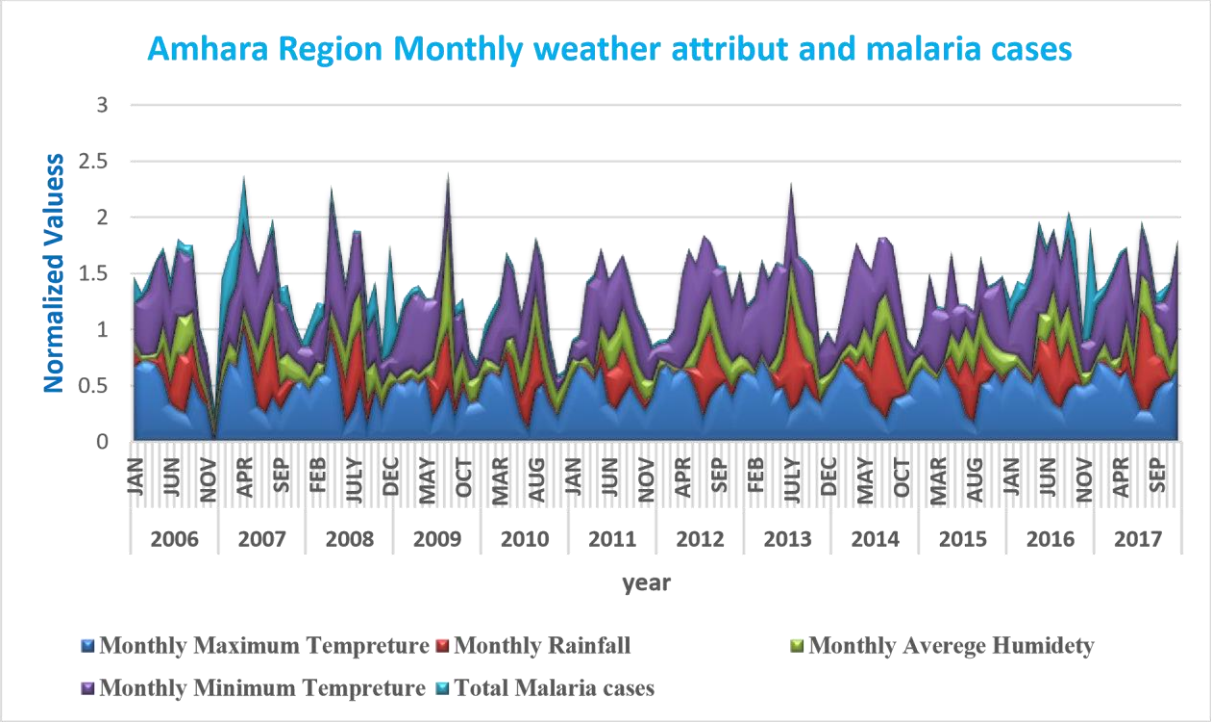


Figure 3.14: Amhara Region Monthly Normalized Attributed data in chart (2006-2017

3.4 Model Development

For this study, Long Short-Term Memory (LSTM) models used to build for model prediction.

3.4.1 Recurrent Neural Networks

Recurrent Neural Networks are the networks with loops in them and allowing information to persist.

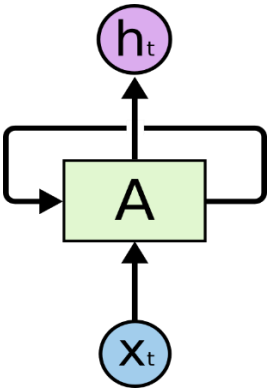


Figure 3.15: Looped Recurrent Neural Networks

In the above diagram, a chunk of a neural network, AA, looks at some input x(t)x(t) and outputs a value h(t)h(t). The loops are allows information to be passed from one step of the network to the next step of the network, that makes recurrent neural networks seem kind of mysterious. The RNN can be thought of as multiple copies of the same network, each passing a message to a successor (A. Sherstinsky, 2020).

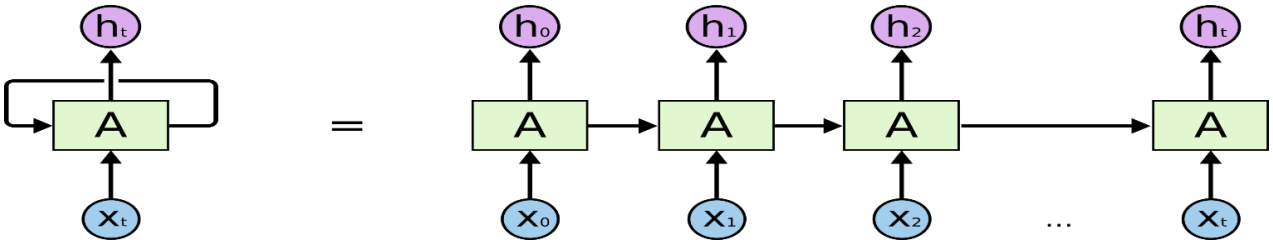
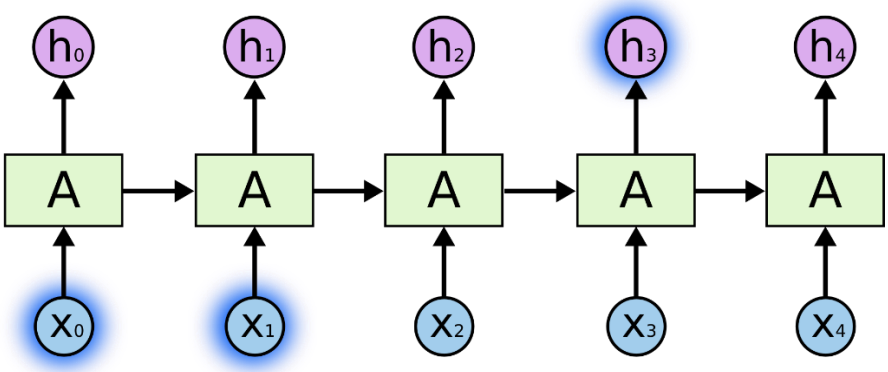


Figure 3.16: An unrolled recurrent neural network

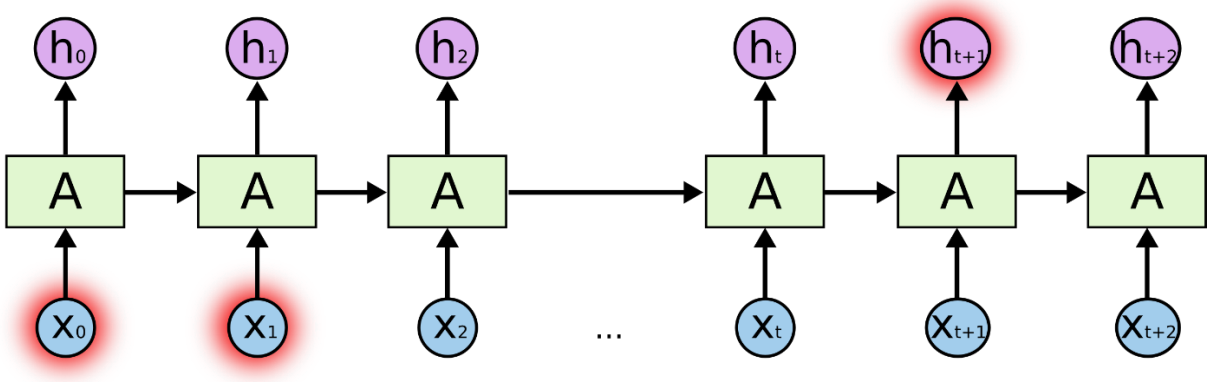
The above figure chain-like nature reveals that RNN are intimately related to a sequences and lists (H. Apaydin, H. Feizi, M. T. Sattari, and M. S. Colak,). They are also the natural architecture of the neural network to use for such data.

3.4.2 The Problem of Long-Term Dependencies

RNN might be able to connect previous information to the present task. If RNNs could do this, they'd be extremely useful (J. Brownlee). Sometimes, we only need to look at recent information to perform the present task.



RNNs become unable to learn to connect the information.



3.4.3 LSTM Networks and their Architecture

In the mid-90s, a variation of the recurrent network with so-called Long Short-Term Memory (LSTM) units was proposed by the German researchers Sepp Hochreiter and Juergen Schmidhuber as a solution to the vanishing gradient problem (J. Brownlee,). LSTMs help preserves the error that can be back propagated through time and layers (T. N. Sainath, O. Vinyals, A. Senior, and N. York,). By maintaining a more constant error, recurrent networks allow continue to learn over many time steps ,there by opening a channel to link causes and effects remotely, this are challenges to machine learning and AI (M. Luong,2016). An LSTM contains information outside the normal flow of the recurrent network in a gated cell. Information can be stored in, written to, or read from a cell(Y. Yu, X. Si, C. Hu, and J. Zhang,2019).

The cell makes decisions about what to store, and when to allow reads, writes, and erasures, via gates that open and close. These gates are analog, implemented with element-wise multiplication by a sigmoid, which are in the range of 0 and 1. Gates act on the signals as they receive, and similar to the neural network's nodes, they can block or pass information based on its strength and import, which they filter with their own sets of weights. The weights are like the weights that modulate input and hidden states, are adjusted via the recurrent networks learning process. These are the cells learn when to allow data to enter, leave, or deleted through the iterative process of making guesses, back propagating error, and adjusting weights via gradient descent. So LSTM is a type of RNN architecture used in the field of deep learning (R. C. Staudemeyer and E. R. Morris,,

2019). LSTM networks are well-suited to classifying, processing, and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series(C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, 2018).

An LSTM allows the preservation of gradients. The memory cell remembers the first input as long as the forget gate is open and the input gate is closed. LSTM is considered to avoid the problem of vanishing gradient in RNN **(**D. Monner and J. A. Reggia,, 2012**)**. We can say that when we move from RNN to LSTM (Long Short-Term Memory) we are introducing more & more controlling knobs, which control the flow and mixing of inputs as per trained weights. So, LSTM gives us the most control-ability and thus, better results (A. Elsaid, B. Wild, J. Higgins, and T. Desell, 2016). We have been using stateful long short term memory for my automated real-time prediction, as i need the model to transfer states between batches.

An LSTM network is a type of recurrent neural network that can learn long-term dependencies between time steps of sequence data. Long Short Term Memory networks usually called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies that explicitly designed to avoid the long-term dependency problem (A. Elsaid, B. Wild, J. Higgins, and T. Desell, 2016). Remembering information for long periods of time is practically their default behavior, not something they struggle to learn. All recurrent neural networks have the form of a chain of repeating modules of the neural network (F. A. Gers, J. Schmidhuber, and F. Cummins, 2000). In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.

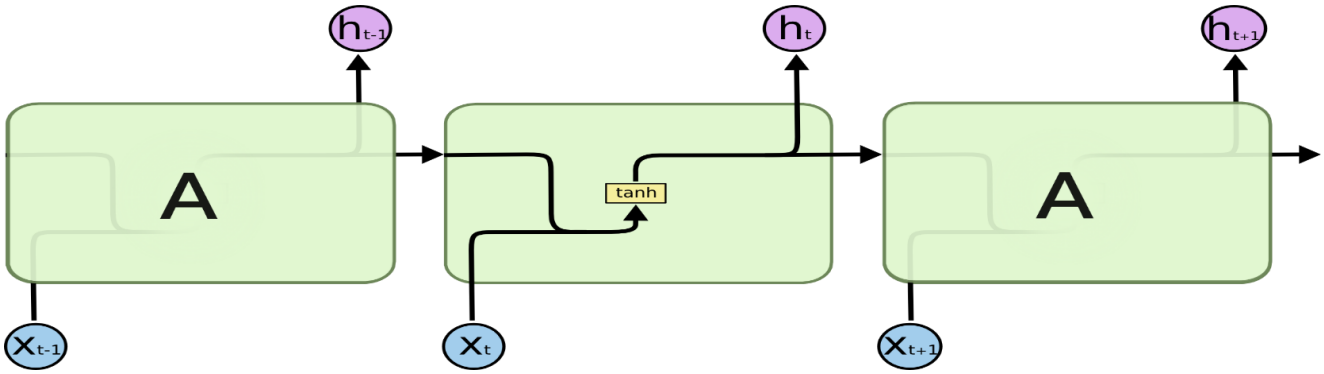


Figure 3.17: The repeating module in a standard RNN contains a single layer

LSTMs also have this chain-like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.

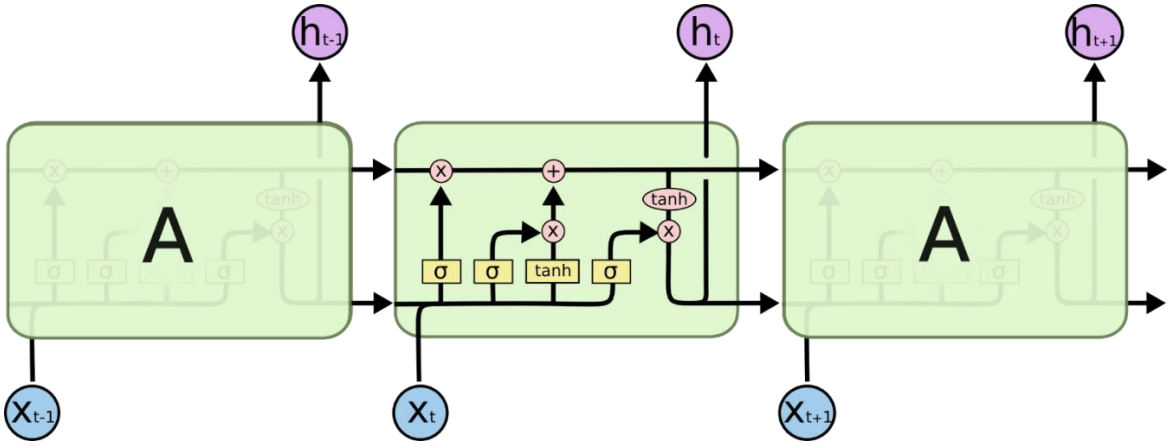


Figure 3.18: The repeating module in an LSTM contains four interacting layers.

The core components of an LSTM network are a sequence of the input layer and LSTM layer(F. A. Gers and N. N. Schraudolph, 2002). A sequence input layer inputs sequence or time-series data into the network. LSTM layer learns long-term dependencies between time steps of sequence data. So we can conclude that it is a recurrent neural network that is trained using back propagation through time and overcomes the vanishing gradient problem(F. A. Gers and F. Cummins,1999). As such, it can be used to create large recurrent networks that in turn can be used to address difficult sequence problems in machine learning and achieve a state of the art results. Instead of neurons, LSTM networks have memory blocks that are connected through layers. A block has components that make it smarter than a classical neuron and a memory for recent sequences, that contains gates that manage the block's state and output (F. A. Gers and F. Cummins,1999). A block operates upon an input sequence and each gate within a block uses the sigmoid activation units to control whether they are triggered or not, making the change of state and addition of information flowing through the block conditional.

There are three types of gates within a unit**:**

Forget Gate: conditionally decides what information to throw away from the block.

Input Gate**:** conditionally decides which values from the input to update the memory state.

Output Gate**:** conditionally decides what to output based on input and the memory of the block.

Each unit is like a mini-state machine where the gates of the units have weights that are learned during the training procedure. The LSTM does have the ability to remove or add information to the cell state, carefully regulated by

structures called gates (F. A. Gers and N. N. Schraudolph,,2002). Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a point wise multiplication operation. The sigmoid layer outputs numbers between zero and one, describing how much of each component should be let through (D. Monner and J. A. Reggia,2012). A value of zero means “let nothing through,” while a value of one means “let everything through”

3.4.4 LSTM working principles

Long short-term memory (LSTM) has the ability to learn long-term dependencies and works tremendously well on a large variety of problems. It is generally designed in a chain like structure, looping back to the previous layers. LSTM’s are augmented by gates called “forget” gates. By controlling these gates, errors can be back propagated through any number of virtual layers..

LSTM networks address the problem of vanishing gradients of RNN by splitting in three inner-cell gates and build so called memory cells to store information in a long range context (Mosavi et al., 2019). As we can see from figure 3.19, a typical LSTM cell is configured mainly by four gates: input gate, input modulation gate, forget gate and output gate. The input gate takes a new input information from outside and process newly coming information. Memory cell input gate takes input from the output of the LSTM cell in the last iteration. Forget gate schedules when to forget the output data and thus selects the optimal time lag for the input sequence. Output gate takes all results processed and generate the output.

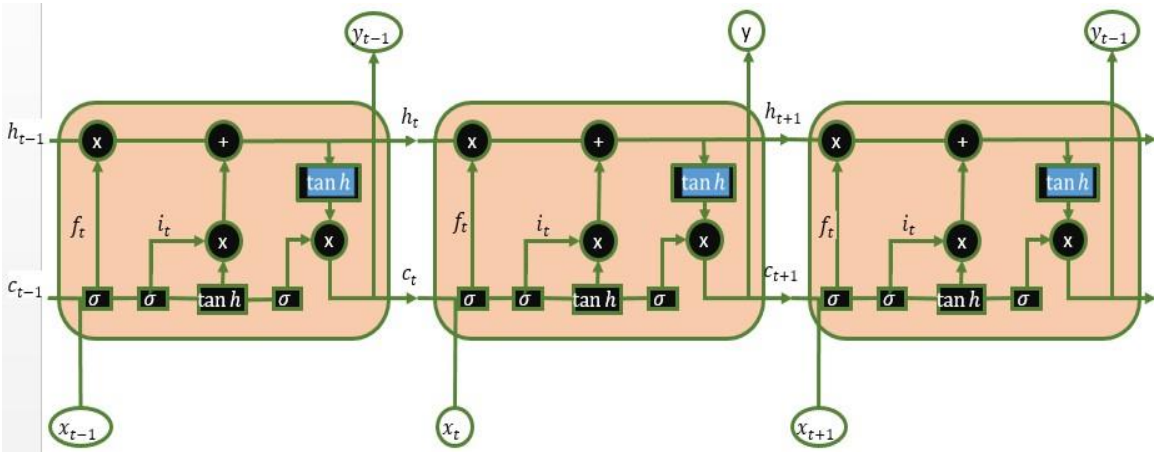


Figure 3.19: Typical LSTM network

The Cell state and the hidden state are the two states that are being transferred to the next cell. Then the LSTM memory is responsible for capturing things and manipulations to this memory. The manipulation of the captured information is done through three major mechanisms, called gates**.** Each of them is being discussed below.

A. Forget gate

A forget gate is responsible for forgetting or removing information from the cell state. The information that is not important for the model to understand things or the data that is of less importance is removed via multiplication of a filter. This helps for optimizing the prediction performance of the LSTM network.

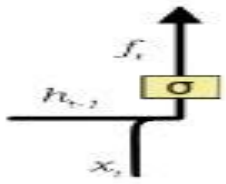


Figure 3.20: Forget gate of LSTM cell

This forget gate takes h_{t-1} and x_t as an inputs variables. Where h_{t-1} is the hidden state from the previous cell states or the output of the previous cell states and x_t is the input at that particular time period. Then, the given inputs are multiplied by the weight matrices and a bias is added to the result. Following this, the sigmoid function is applied to the input value. The sigmoid function outputs a vector, with values ranging from 0 to 1, based on each number in the cell state. It is responsible for deciding which values to keep and which to discard. For instance, if the output for a particular value in the cell state is '0', it means that the forget gate wants the cell state to forget that piece of information completely from the memory. Similarly, a '1' means that the forget gate wants to remember the entire piece of information and this vector output from the sigmoid function is multiplied to the cell state.

B. Input gate

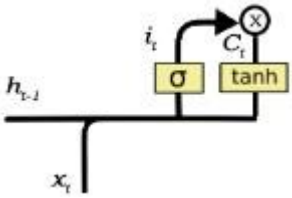


Figure 3.21: Input gate of LSTM network

The addition of information to the cell state is the responsibility of the input gate and this addition of information is three-step process as seen from the diagram above. First step is regulating the values need to be added to the cell state using a sigmoid function. This is also very similar to the forget gate and acts as a filter for all the information from h_{t-1} and . The second step is creating a vector containing all possible information that can be added (as perceived from h_{t-1} and x_t) to the cell state using $\tan h$, which outputs values from -1 to +1. Multiplying the value of the regulatory to the created vector (using the $\tan h$ function) and adding this useful information to the cell state via addition operation is the last step.

C. Output gate

The selection of useful information from the current cell state and showing it out as an output is done via the output gate of the model. Information selection via output gate has again three steps: the first step is creating a vector after applying $\tan h$ function to the cell state, thereby scaling the values to the range -1 to +1. The Second is making a filter using the values of h_{t-1} and x_t, such that it can regulate the values that need to be output from the vector created above using a sigmoid function. The last step is multiplying the values of this regulatory filter to the vector created in step 1, and finally sending it out as an output and also to the hidden state of the next cell state. Here is its structure:

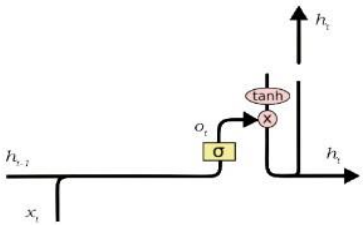


Figure 3.22: Output gate of LSTM network

3.5 Setup the dataset

Dividing the available sample dataset into a training dataset and the testing dataset is the most common approach used in machine learning to build and evaluate prediction and classification models. In this work the entire dataset is divided into three parts, those are: training dataset, validating and testing dataset which are approximately 70% for training 15 % for validating and 15% for testing which means that 8 years (2006 to 2013) data are used for training the LSTM model 2 years (2014 to 2015) used for validating the trained model, and two years (2016-2017) for testing the developed model. Those datasets are shown below

Table 3.10: Amhara region monthly training dataset (2006-2007)

year	month	T max	Rainfall	Humidity	Tmin	Malaria cases
2006	JAN	0.657142857	0.08142	0.083619866	0.32	0.249606794
2006	FEB	0.714285714	0	0.05604347	0.5	0.059574764
2006	MAR	0.714285714	0.018182	0.035682185	0.59	0.100666811
2006	APR	0.685714286	0.046801	0.053737284	0.82	0.018486813
2006	MAY	0.571428571	0.256902	0.210526316	0.64	0.059574764
2006	JUN	0.342857143	0.166162	0.263157895	0.57	0.121207715
2006	JUL	0.285714286	0.490741	0.338288796	0.59	0.102919553
2006	AUG	0.257142857	0.513973	0.336576736	0.56	0.08011874
2006	SEP	0.571428571	0.324074	0.26006192	0.48	0.121206691
2006	OCT	0.4	0.147138	0.160876575	0.02	0.059574764
2006	NOV	0.314285714	0.050337	0.101503759	0.27	0.039028741
2006	DEC	0	0.058923	0.095390278	0	0.141750667
2007	JAN	0.485714286	0.008081	0.152316275	0.25	0.561305279
2007	FEB	0.714285714	0	0.175222721	0.35	0.458907953
2007	MAR	0.657142857	0.026599	0.113580916	0.59	0.418928965
2007	APR	1	0.075926	0.120101725	0.73	0.46018792
2007	MAY	0.685714286	0.171044	0.142043914	0.73	0.01882882
2007	JUN	0.314285714	0.273064	0.242096112	0.64	0.024154505
2007	JUL	0.257142857	0.573569	0.323094263	0.57	0.015955551
2007	AUG	0.4	0.597811	0.326090368	0.55	0.120996777
2007	SEP	0.285714286	0.213131	0.23772667	0.5	0.134291021
2007	OCT	0.4	0.161953	0.223580774	0.41	0.203208518
2007	NOV	0.514285714	0.017845	0.183768244	0.3	0.058343949
2007	DEC	0.542857143	0	0.123639269	0.18	0.045015913

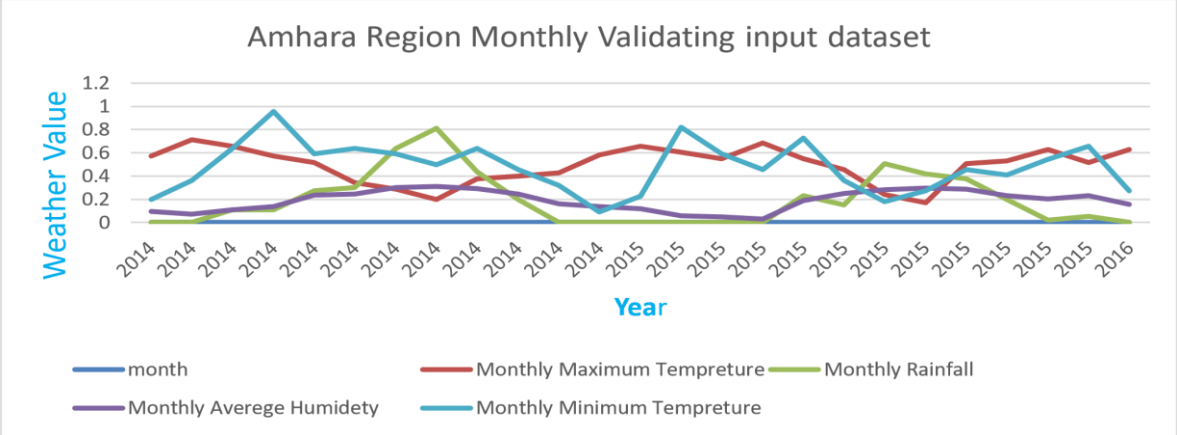


Figure 3.23: Amhara region monthly validating input dataset

Table 3.11: Amhara region monthly validating dataset (2014-2015)

year	month	Tmax	Rainfall	Humidity	Tmin	Malaria cases
2014	JAN	0.571429	0	0.096246	0.2	0.006613
2014	FEB	0.714286	0	0.074051	0.363636	0.005558
2014	MAR	0.657143	0.110943	0.108017	0.636364	0.004944
2014	APR	0.571429	0.112121	0.137771	0.954545	0.001937
2014	MAY	0.514286	0.275589	0.236207	0.590909	0.002046
2014	JUN	0.342857	0.300337	0.246351	0.636364	0.001053
2014	JUL	0.285714	0.637037	0.302764	0.590909	0.000879
2014	AUG	0.2	0.809428	0.311752	0.5	0.001017
2014	SEP	0.377143	0.43771	0.292127	0.636364	0.005246
2014	OCT	0.4	0.197643	0.245495	0.454545	0.007196
2014	NOV	0.428571	0	0.161477	0.318182	0.006582
2014	DEC	0.582857	0.000673	0.139476	0.090909	0.019207
2015	JAN	0.658286	0	0.121371	0.227273	0.034892
2015	FEB	0.605714	0.001347	0.06064	0.818182	0.020477
2015	MAR	0.547429	0.000673	0.049593	0.590909	0.020044
2015	APR	0.685714	0	0.032065	0.454545	0.018219
2015	MAY	0.548571	0.230303	0.190624	0.727273	0.009248
2015	JUN	0.457143	0.150337	0.250332	0.363636	0.008437
2015	JUL	0.24	0.508754	0.283503	0.181818	0.011296
2015	AUG	0.171429	0.419024	0.296771	0.272727	0.010279
2015	SEP	0.508571	0.376936	0.288147	0.454545	0.028702
2015	OCT	0.528571	0.196465	0.230429	0.409091	0.022568
2015	NOV	0.628571	0.020539	0.203892	0.545455	0.020515
2015	DEC	0.514286	0.053535	0.230429	0.654545	0.025351

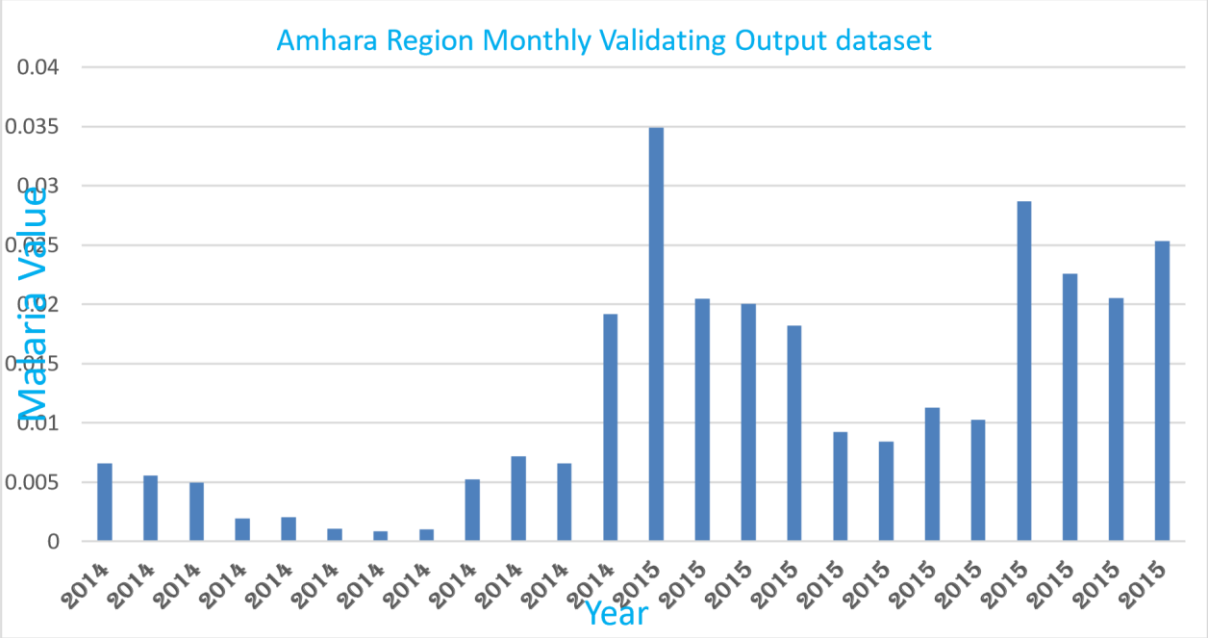


Figure 3.24: Amhara region monthly validating output dataset

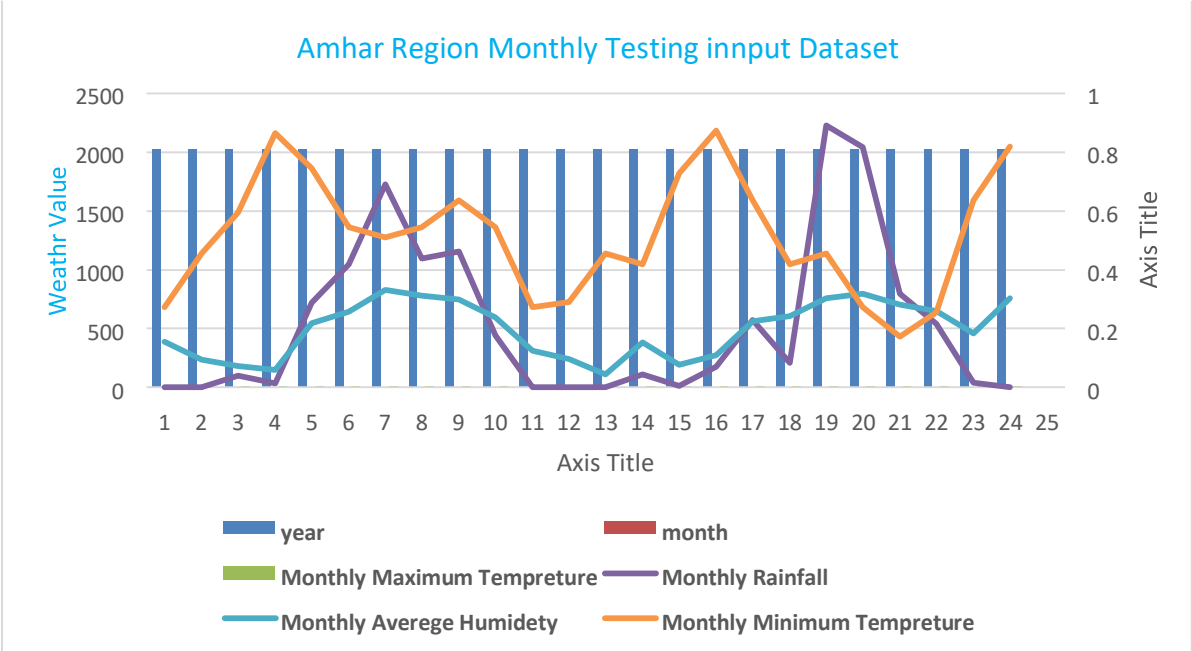


Figure 3.25: Amhara region monthly testing input dataset

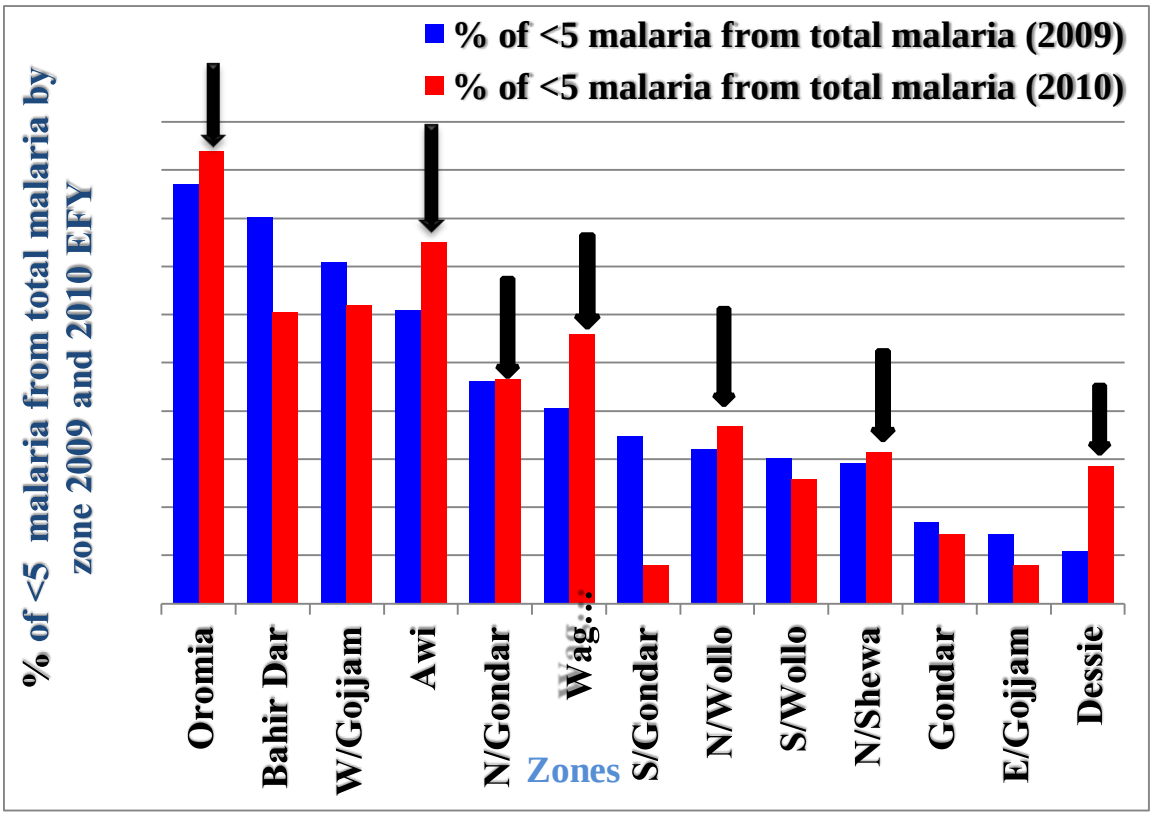


Figure 3.26: Amhara region % of < 5-year Malaria from total Malaria by zone 2009 and 2010 EFY

3.6. Error metrics

In this study, mean squared error metrics is used to calculate the error for the model, MSE is one of the most commonly used measures of success for numeric prediction(Mandale & jadhawar,2015).[52] This value is computed by taking the average of the squared difference between each computed value and its corresponding actual value. MSE Measures using the following values.

$$\mathbf{MSE}=\mathbf{1}/n \sum_{k=0}^{n}(\mathbf{actual}-\mathbf{predicted})^2$$

Where n is the number of raw data

3.7. Model Implementation

This work implemented using Anacoda3 python with different libraries have been used. Some of them among the libraries are pandas, Tensor flow, Keras, NumPy, Matplotlib and different helping tools are imported to anaconda.

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.it presents a diverse range of utilities, ranging from parsing multiple file-formats to converting entire data tables into a Numpy matrix array. This makes pandas trusted in data science and machine learning.

Tensor Flow: is a Python library for high-performance numerical calculations that allows users to create sophisticated deep learning and machine learning applications. Tensor Flow is supported on a number of platforms. Tensor Flow coda packages are available for Windows, Linux, and macOS. so it is a software framework for numerical competitions based on dataflow graphs.

Keras: is the most popular and powerful library running on top of the Tensor flow. It is a powerful and easyto-use free open-source Python library for developing and evaluating deep learning models. It wraps the efficient numerical computation libraries Theano and Tensor Flow and allows you to define and train neural network models in just a few lines of code.

NumPy: are one of the fundamental package in Python used for scientific calculation that helps for inserting any type of mathematical operation in the code.

Matplotlib: are a Python 2D plotting library that is used to plot any type of charts in Python.

The pseudo-code implementation on python using Jupiter notebook editor as follows

Step 1. Import all necessary packages on the python notebook editor

Step 2. Load the dataset to the editor

Step3. Split the data into training, validating and testing data

Step4. Design The model

Step5. Compil the model

Step6. Fit the model with training dataset and validating dataset with iterations (epochs)

Step7. Evaluate the model with the testing dataset

Step8. Predict the future

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1 Evaluating the model

The Evaluation of the ANN model was done using simulation experimental methods. The tests were done using the developed model, based on a configuration of with 70% data for training (January 2006 to December 2013), 15% data for validating (January 2014 to December 2015) and also 15 % are used for testing the data that are from January 2016 to December 2017. The training data are used for developing the model, testing data are used for evaluating the developed model, and validation data are used for tuning the parameters of a model.

4.2 Result

This was done using the testing dataset. The tables and graphs show the actual and predicted data for the year 2016 -2017

Table 4.1: Amhara Region Actual and predicted Malaria cases of the year 2016

Months	Actual	Predicted	Error
JAN	0.210373259	0.276471	0.066097741
FEB	0.211453551	0.29101455	-0.07956099
MAR	0.121454493	0.46878362	-0.34732817
APR	0.109072608	0.7100166	-0.60094399
MAY	0.08870066	0.76369584	-0.67499518
JUN	0.060181981	0.6890776	-0.62889561
JULY	0.28991755	0.12665607	0.16326148
AUG	0.021111257	0.41047505	-0.3893638
SEP	0.605197074	0.5074938	0.097703274
OCT	0.311173187	0.68205684	-0.37088365
NOV	0.101329322	0.42487	-0.32354067
DEC	1	0.34972432	0.65027568

The result can be represented by graphs as shown below

Mse: 0.078503

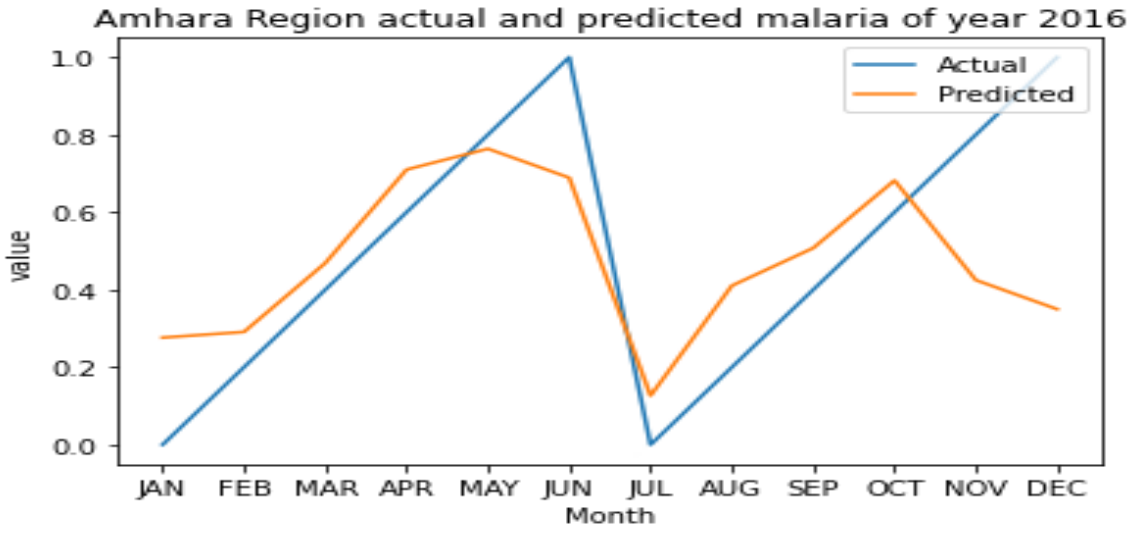


Figure 4.1: Actual and Predicted malaria of the year 2016 using graph in Amhara Region

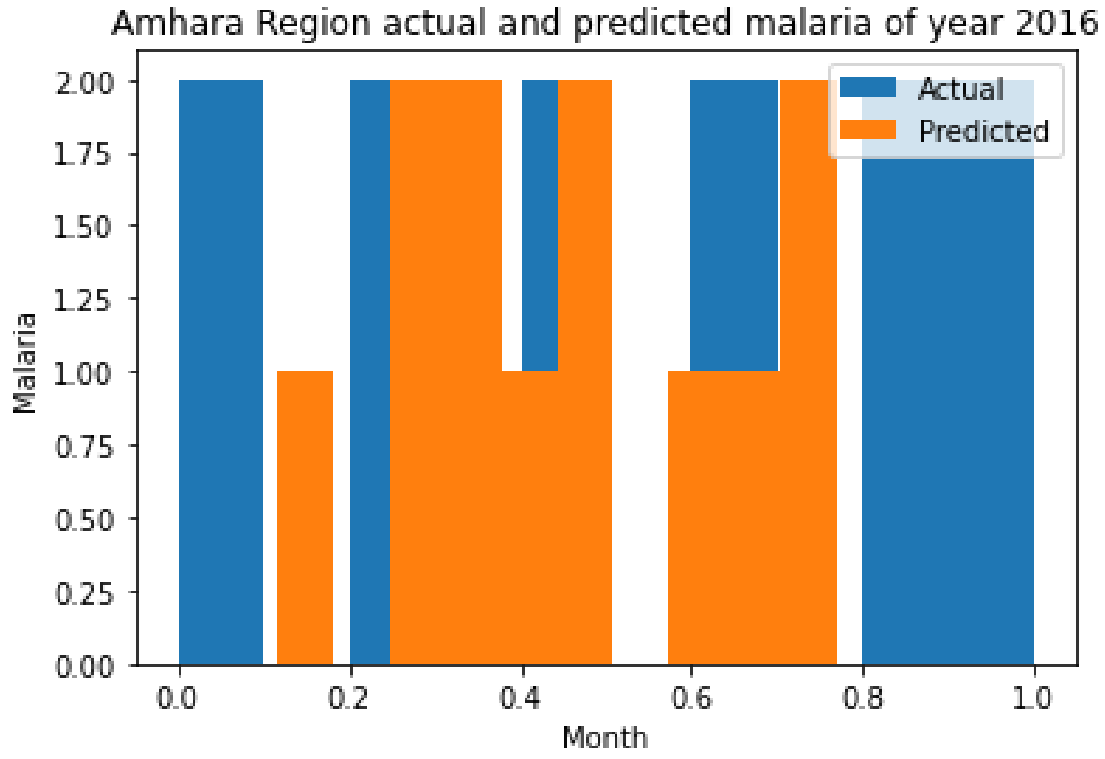


Figure 4.2: Actual and Predicted malaria of the year 2016 using Histogram in Amhara Region

Table 4.2: Amhara region actual and predicted Malaria of The year 2017

Month	Actual	Predicted	Error
JAN	0.121067431	0.02606523	0.095002201
FEB	0.091018935	0.276909	-0.18589006
MAR	0.069893343	0.5063846	-0.43649125
APR	0.066532663	0.6876628	-0.62113013
MAY	0.018912786	0.01894173	0
JUN	0.016268887	0.02776064	-0.01149175
JULY	0.035817561	0.13395071	-0.09813315
AUG	0.060388823	0.11282229	-0.05243346
SEP	0.091108021	0.07308913	0.018018891
OCT	0.12178114	0.039884442	0.081896698
NOV	0.041862075	0.5749183	-0.53305622
DEC	0.055070306	0.66055113	-0.60548082

The Result can be represented using the graph as shown below:

Mse: 0.113605

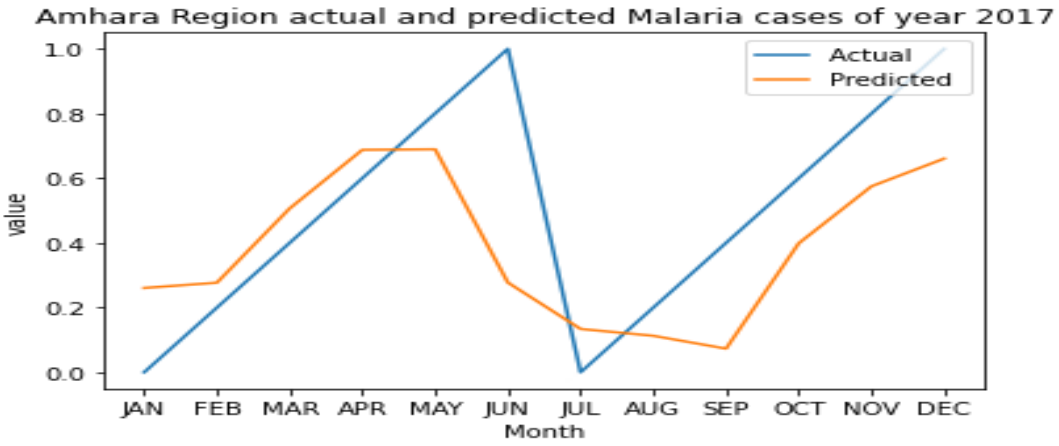


Figure 4.3: Actual and predicted malaria of the year 2017 using graph in Amhara Region

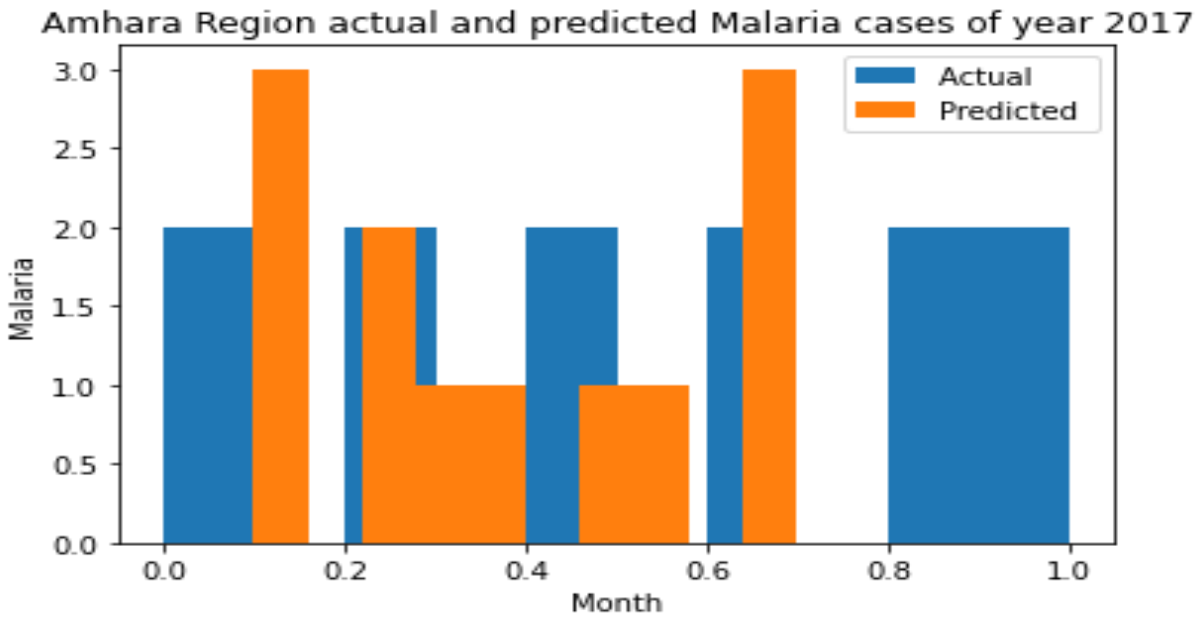


Figure 4.4: Actual and predicted malaria of the year 2017 using histogram in Amhara Region

Mse: 0.096054

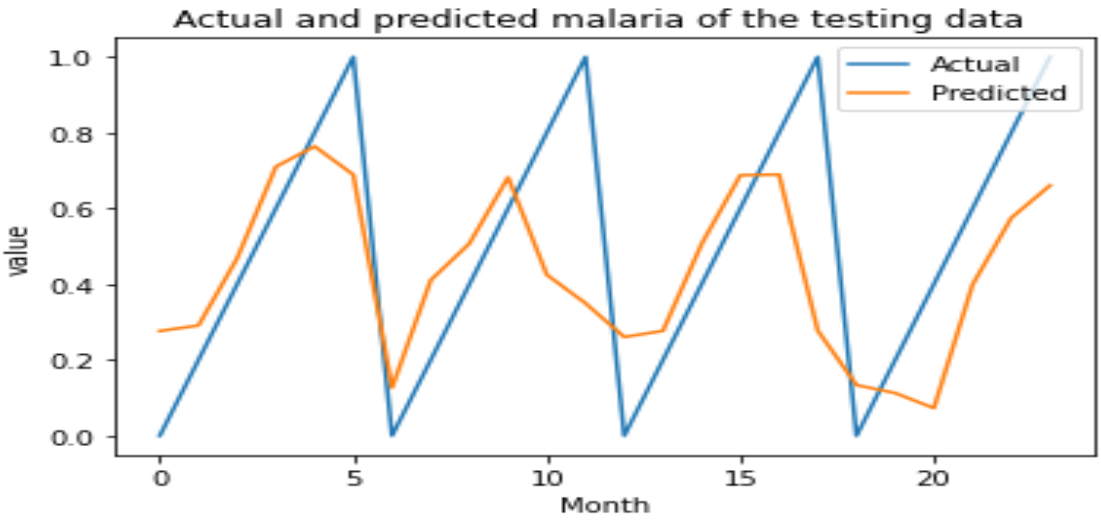


Figure 4.5: Actual and predicted malaria of the testing data-using graph in Amhara region

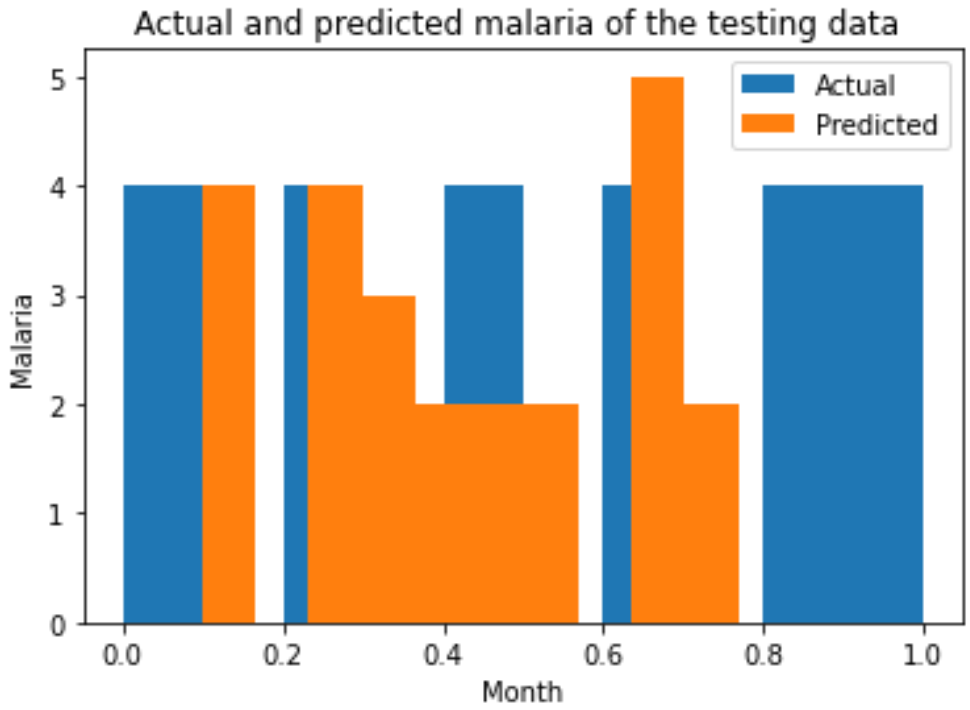


Figure 4.6: Actual and predicted malaria of the testing data using Histogram in Amhara region

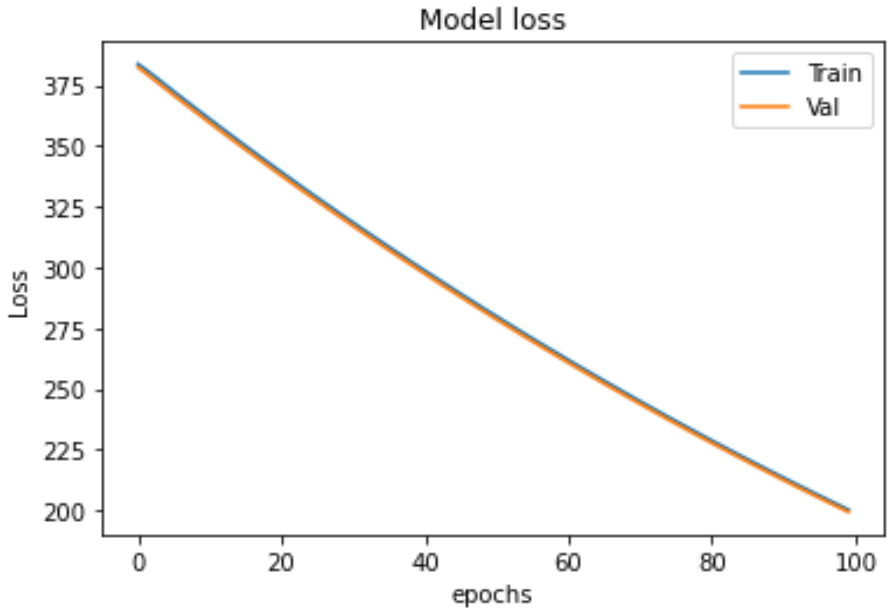


Figure 4.7: Model Loss of the developed model for Amhara Region

Table 4.3: Actual and predicted Malaria of 2016 in W/Go jam and S/Gondar Zone

	West Go jam Zone			South Gondar Zone		
Month	Actual	predicted	Error	Actual	Predicted	Error
JAN	0.031244207	0.59238553	0.561141323	0	0.4010156	0.401016
FEB	0.036763949	0.25812745	0.221363501	0.00640947	0.6889102	0.682501
MAR	0.048812638	0.40220404	0.353391402	0.010334887	0.68505806	0.674723
APR	0.060130167	0.543797	0.483666833	0.041523553	0.9357112	0.894188
MAY	0.060140465	0.9109047	0.850764235	0.053269136	0.833201	0.779932
JUN	0.04139806	0.8885162	0.84711814	0.096172718	0.0384276	-0.05775
JULY	0.04150104	0.05399678	0.01249574	0.010334887	0.08388602	0.073551
AUG	0.051284164	0.46113926	0.409855096	0.013432287	0.55557615	0.542144
SEP	0.061510102	0.6809664	0.619456298	0.007390824	0.7143739	0.706983
OCT	0.062787057	0.8779956	0.815208543	0.013370952	0.8385512	0.82518
NOV	0.082559265	0.41566294	0.333103675	0.10233685	0.66816103	0.565824
DEC	0.001410829	0.12579988	0.124389051	0.141284347	0.5273048	0.38602

Mse: 0.049267

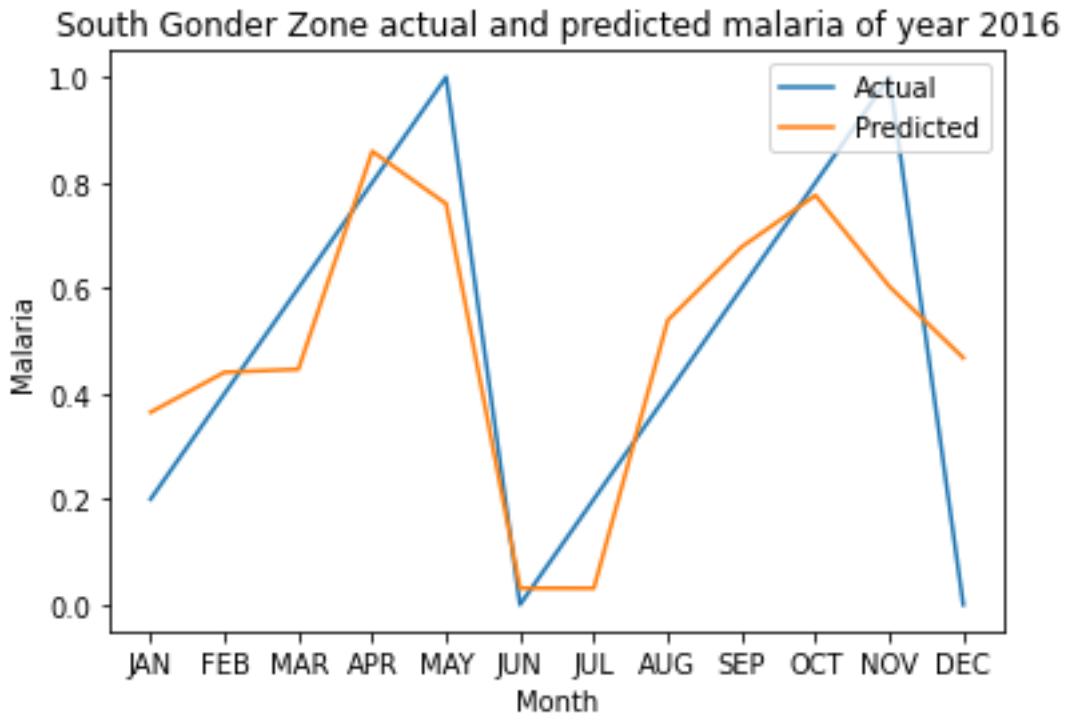


Figure 4.8: Actual and predicted malaria of the year 2016 using graph in S/Gondar

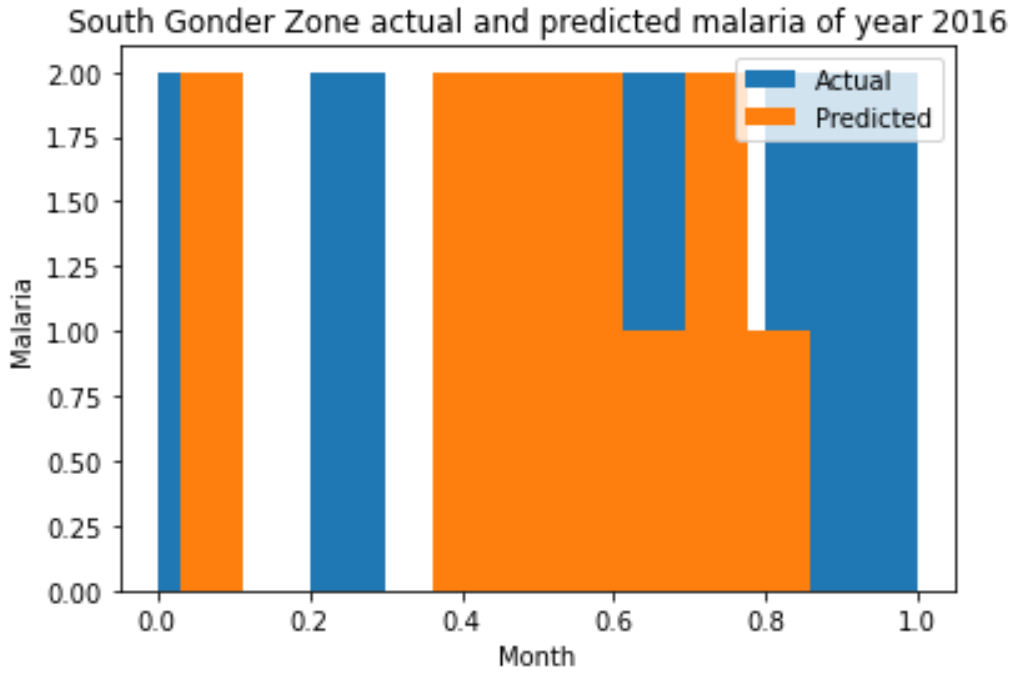


Figure 4.9: Actual and predicted malaria of the year 2016 using Histogram in S/Gondar **Mse: 0.100292**

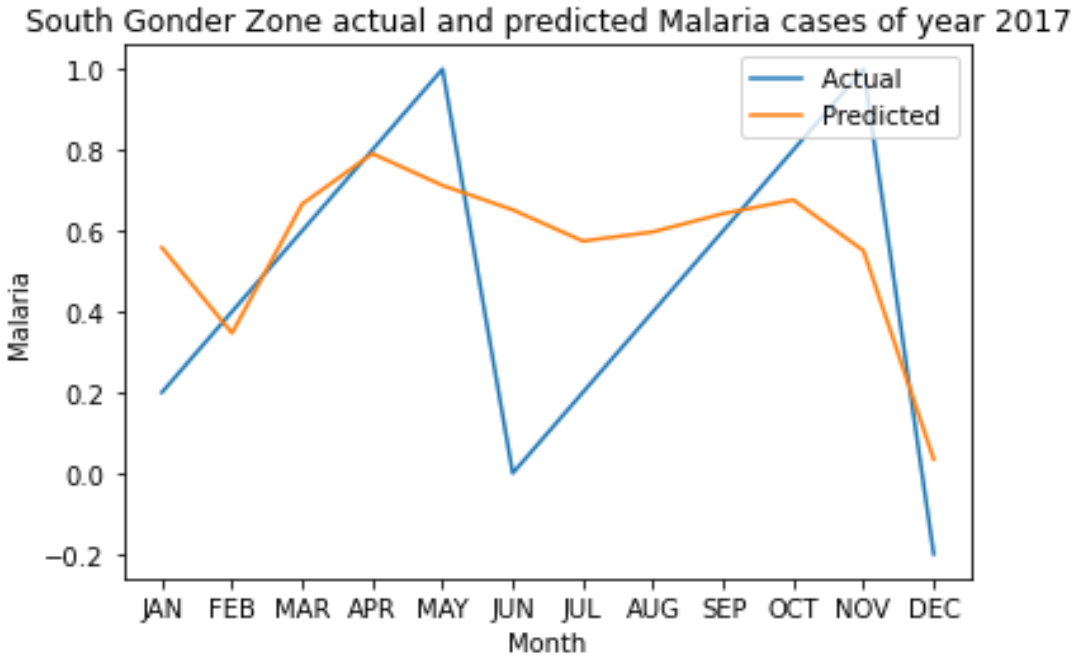


Figure 4.10: Actual and predicted malaria of the year 2017 using graph in S/Gondar

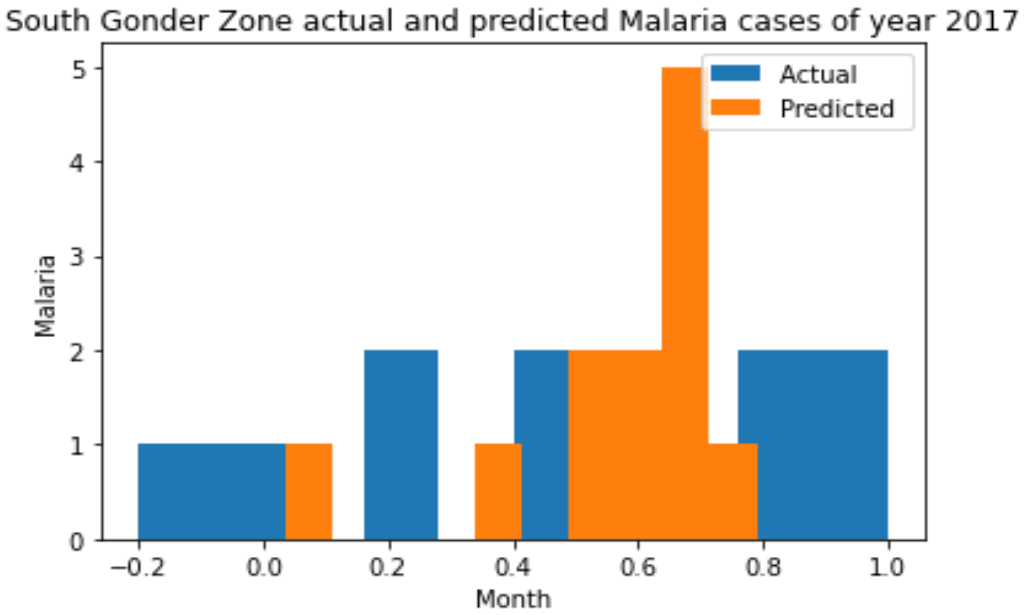


Figure 4.11: Actual and predicted malaria of the year 2017 using Histogram in S/Gondar

Mse: 0.074779

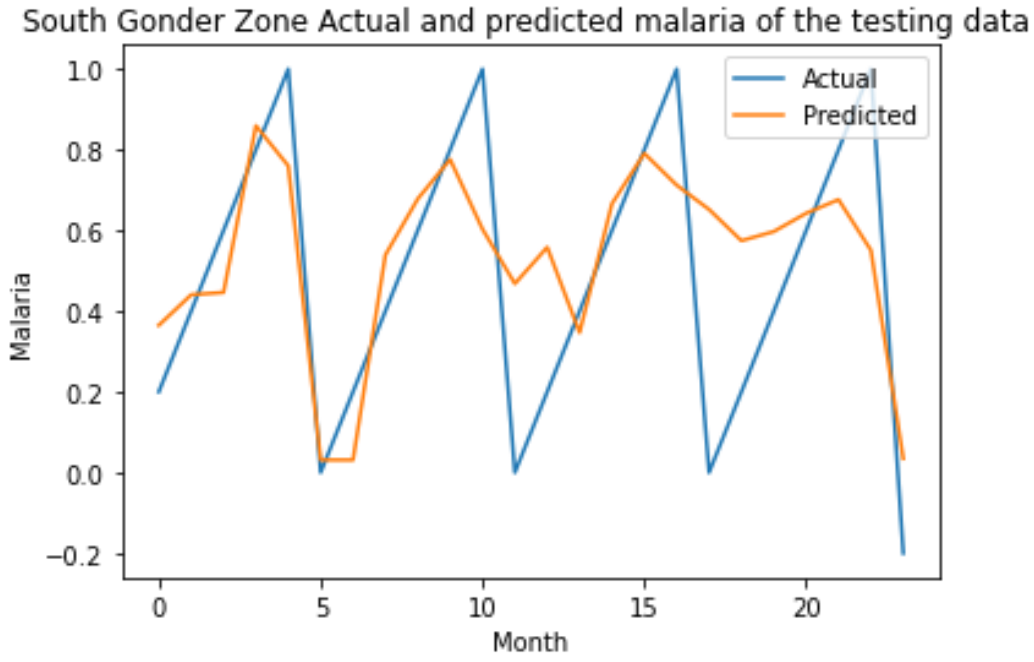


Figure 4.12: Actual and predicted malaria of the testing data using graph in S/Gondar

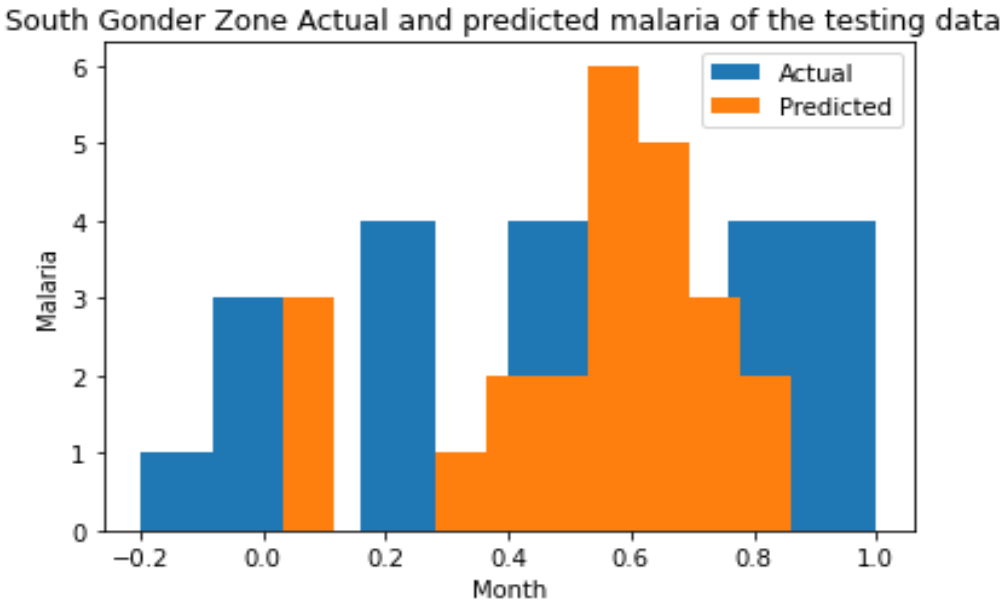


Figure 4.13: Actual and predicted malaria of the testing data using Histogram in S/Gondar

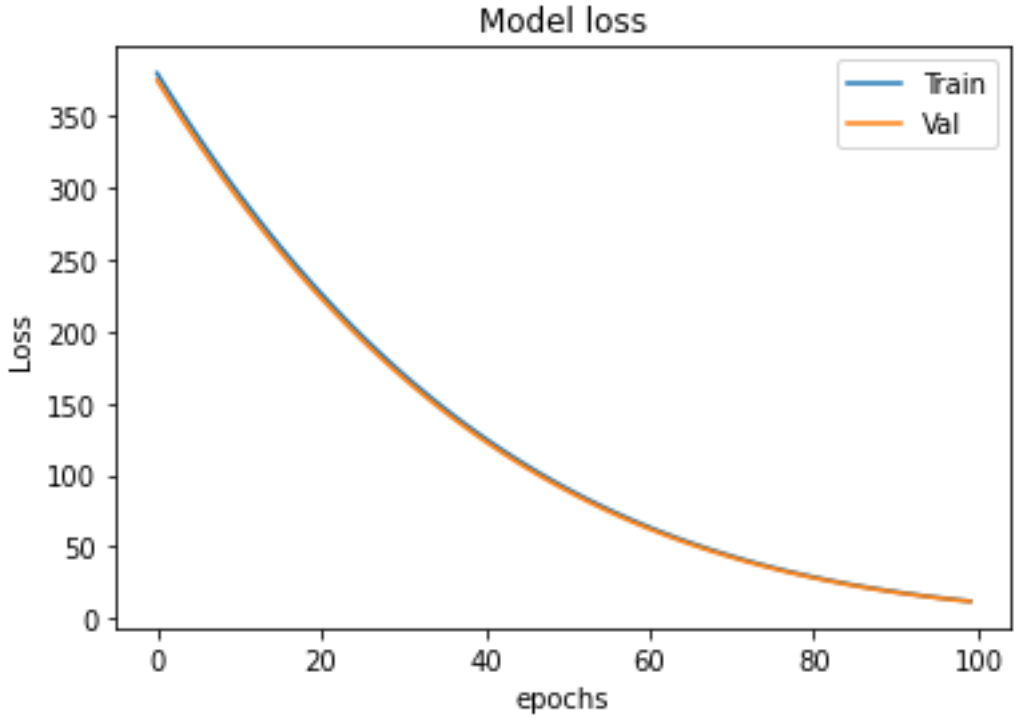


Figure 4.14: Model Loss of the developed model for S/Gondar Region

Model Performance 93.16% and Model Error 6.84%

Table 4.4: Actual and predicted Malaria of 2017 in W/Go jam and S/Gondar zone

	West Go jam zone			**South Gondar zone**		
Month	Actual	predicted	Error	Actual	Predicted	Error
JAN	0.009793422	0.05968494	0.049891518	0.039622179	0.59891915	0.559297
FEB	0.020781414	0.8353118	0.814530386	0.047105005	0.5747778	0.527673
MAR	0.0291949	0.5083878	0.4791929	0.05173577	0.7849566	0.733221
APR	0.031285399	0.88343763	0.852152231	0.071454858	0.8752193	0.803764
MAY	0.031532552	0.9493462	0.917813648	0.080869725	0.7432942	0.662424
JUN	0.01731098	0.92246026	0.90514928	0.110617026	0.5222932	0.411676
JULY	0.026126089	0.00250437	-0.02863045	0.049282385	0.49867716	0.449395
AUG	0.035486994	0.00745704	-0.02802995	0.051797105	0.6041376	0.55234
SEP	0.035517888	0.70054775	0.665029862	0.10233685	0.65805227	0.555715
OCT	0.042716207	0.8786665	0.835950293	0.141284347	0.7289458	0.587661
NOV	0.041923259	0.82857704	0.786653781	0.142204367	0.61607736	0.473873
DEC	-0.009783124	0.4556033	0.445820176	0.145240432	0.03959233	-0.10565

Mse=0.113553

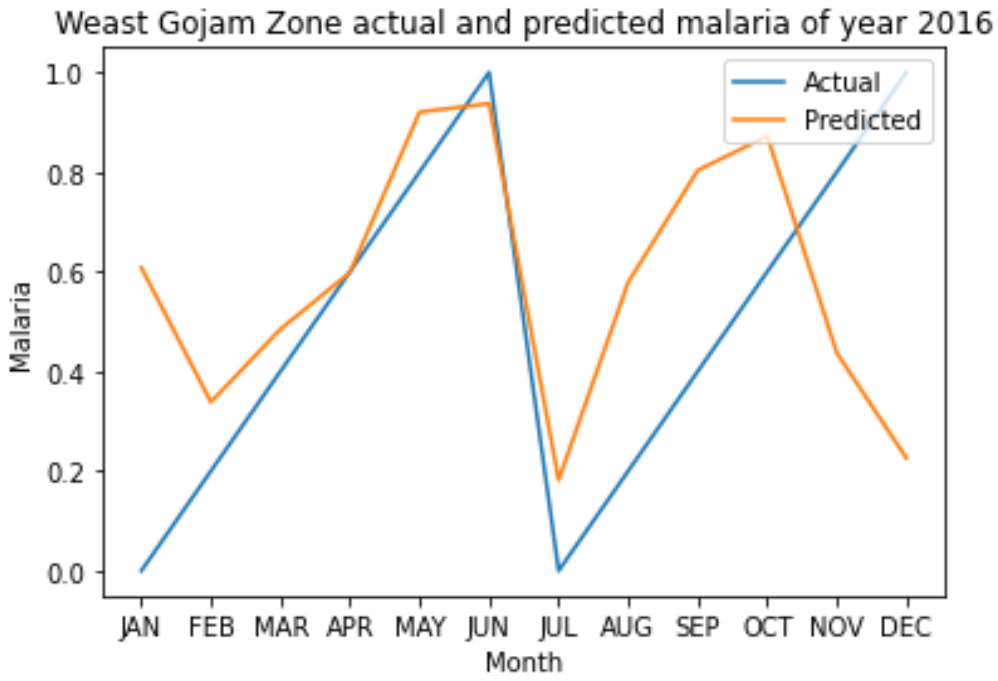


Figure 4.15: Actual and predicted malaria of the year 2016 using graph in W/Gojam

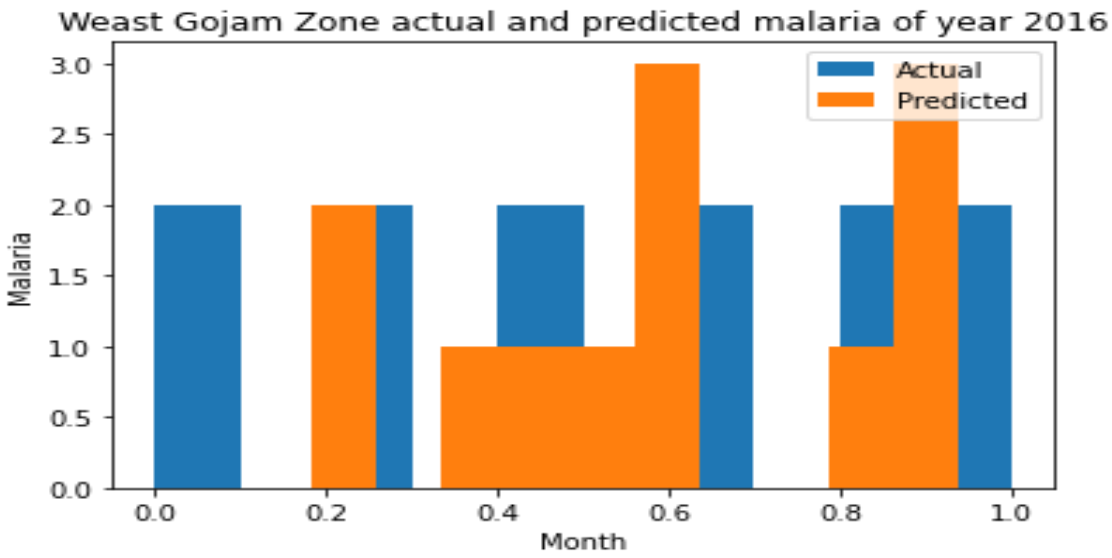


Figure 4.16: Actual and predicted malaria of the year 2016 in W/Gojam Using Histogram

Mse: 0.067914

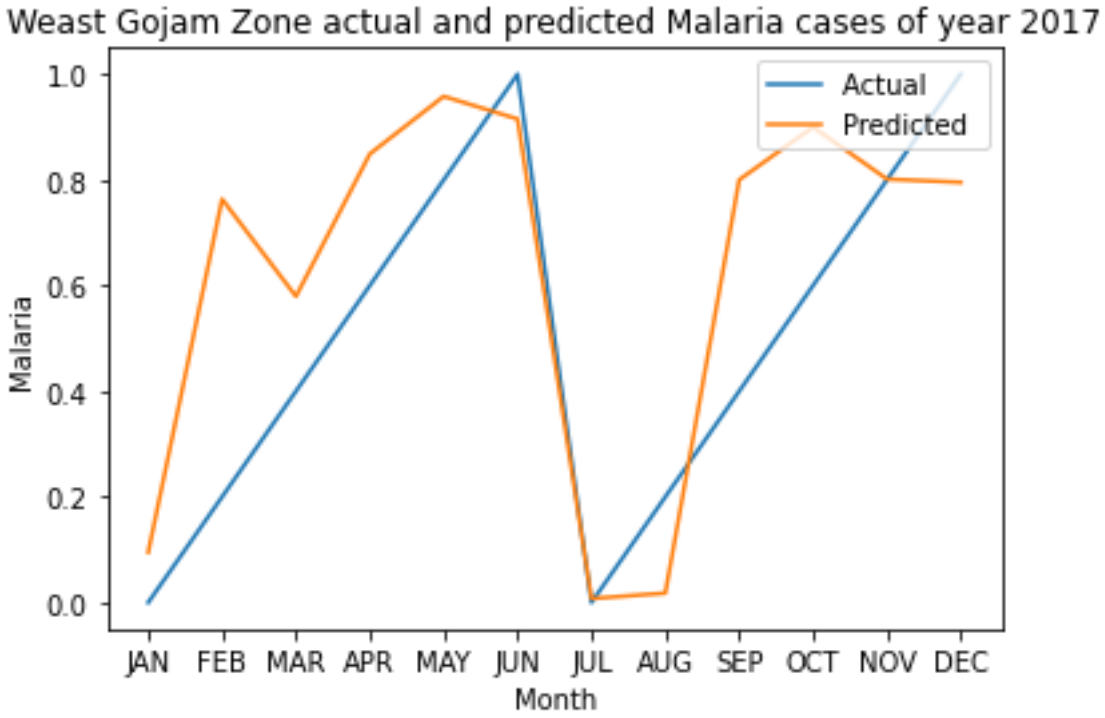


Figure 4.17: Actual and predicted malaria of the year 2017 using graph in W/Gojam

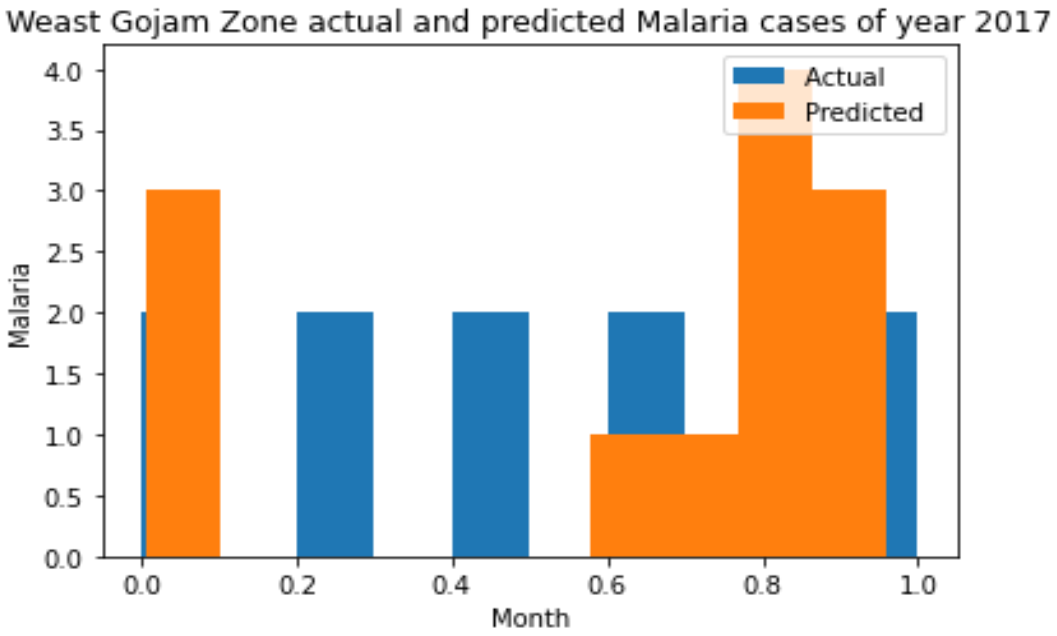


Figure 4.18: Actual and predicted malaria of the year 2017 in W/Gojam Using Histogram

Mse:0.090733

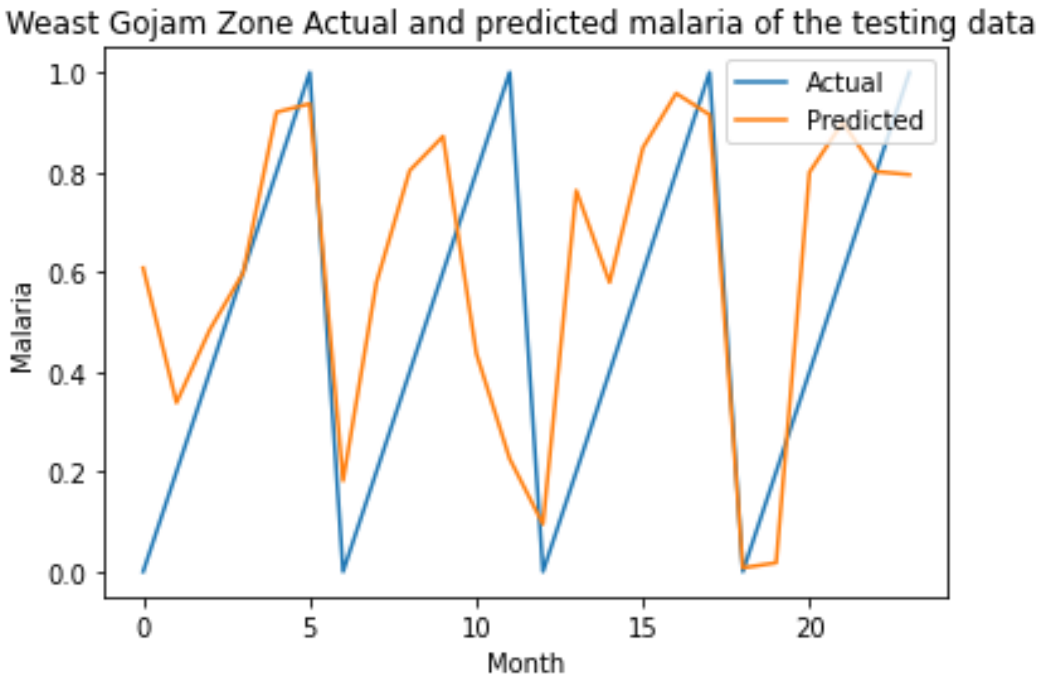


Figure 4.19: Actual and predicted malaria of the testing data-using graph in W/Gojam

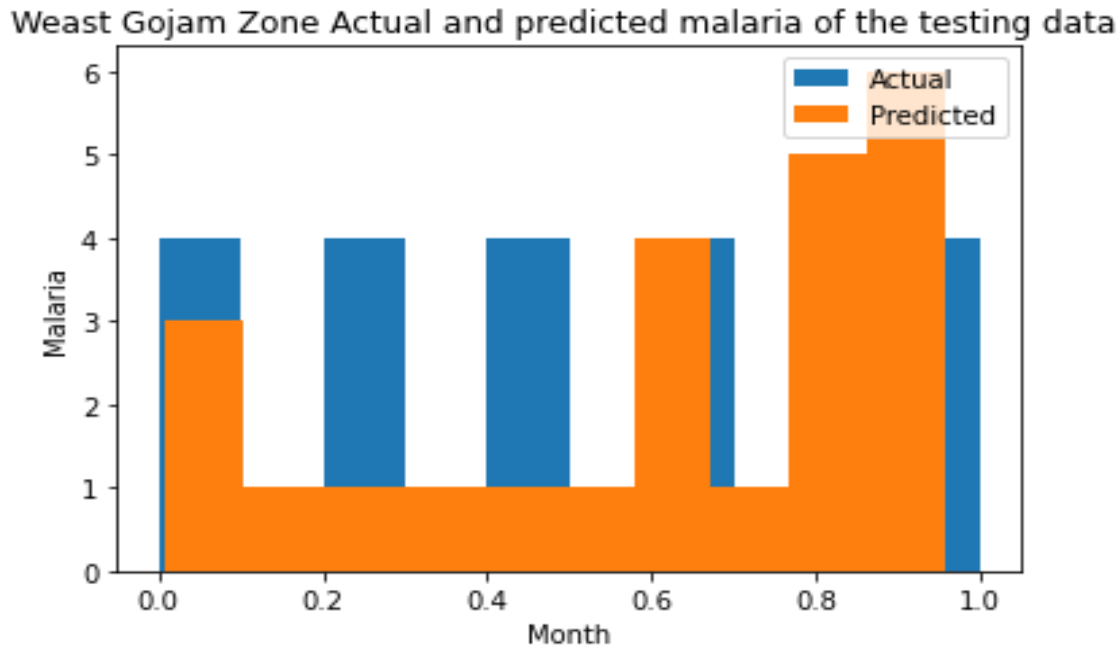


Figure 4.20: Actual and predicted malaria of the testing data using Histogram in W/Gojam

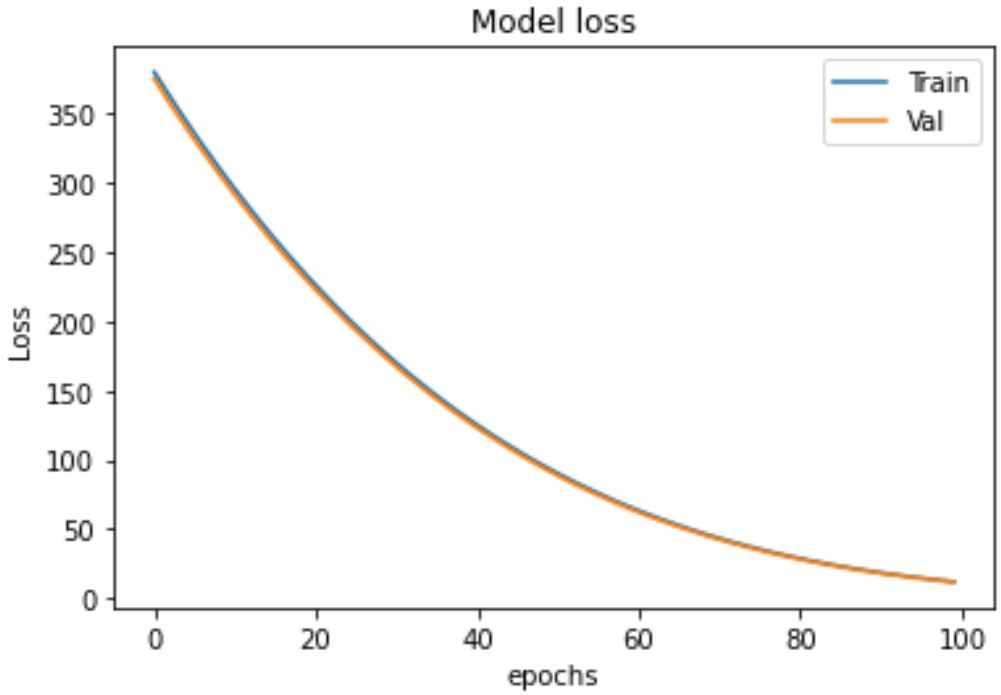


Figure 4.21: Model Loss of the developed model for W/Gojam Zone

Model performance 90.26% and Model Error 9.74%

4.3 Model Diagnostic Plots

The training history of LSTM models can be used to diagnose the behavior of the model [53].

4.3.1 LSTM Under fitted Model Losses

An under fit LSTM model is one that is demonstrated to perform well on the training dataset and weak on the testing dataset. This can be diagnosed from the plot where a training loss is lower than the validation loss, and the validation loss has a trend that suggests further improvements are possible. In this case, performance may be improved by increasing the number of training epochs.

Alternately, a model may be under fit if performance on the training set is better than the validation set and performance has leveled off. An under fit model with insufficient memory cells. In this case, performance may be improved by increasing the capacity of the model, such as the number of memory cells in a hidden layer or the number of hidden layers.

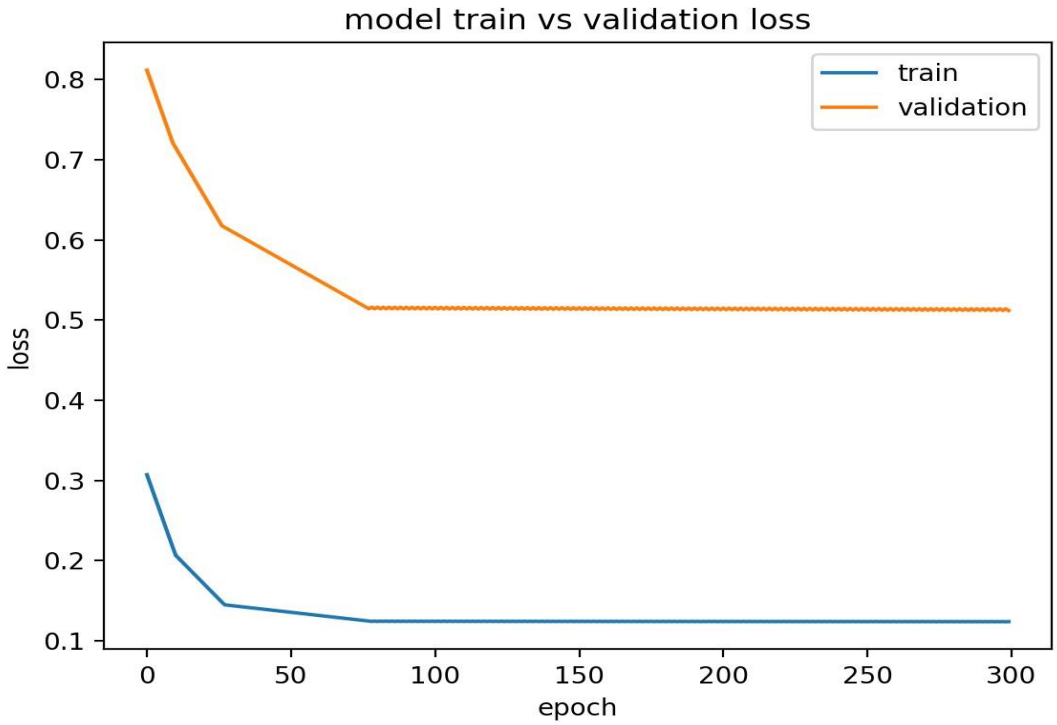


Figure 4.22: Representation of LSTM under fitted model Losses Graphically

4.3.2 LSTM Good Fitted Model

A good fit is a case where the performance of the model is good on both the train and validation sets. This can be diagnosed from a plot where the train and validation loss decrease and stabilize around the same point. Ideally, we would like to see a model performance like this if possible, although this may not be possible on challenging problems with a lot of data. We can conclude that the malaria abundance LSTM model prediction is a good fitted model.

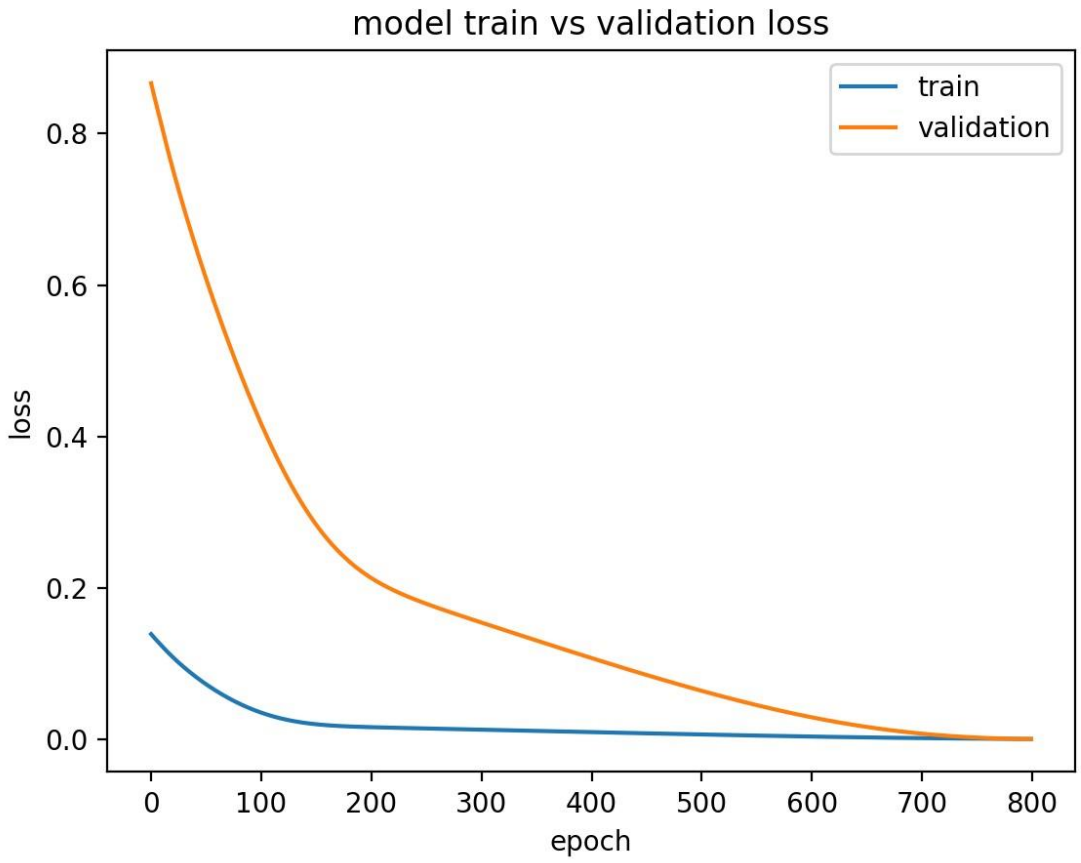


Figure 4.23: Representation of LSTM Good fitted Model Losses Graphically

4.3.3 LSTM Over fitted Model Losses

An over fit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade. This can be diagnosed from a plot where the train loss slopes down and the validation loss slopes down, hits an inflection point and starts to slope up again. In this case, the model training could be stopped at the inflection point. Alternately, the number of training examples could be increased.

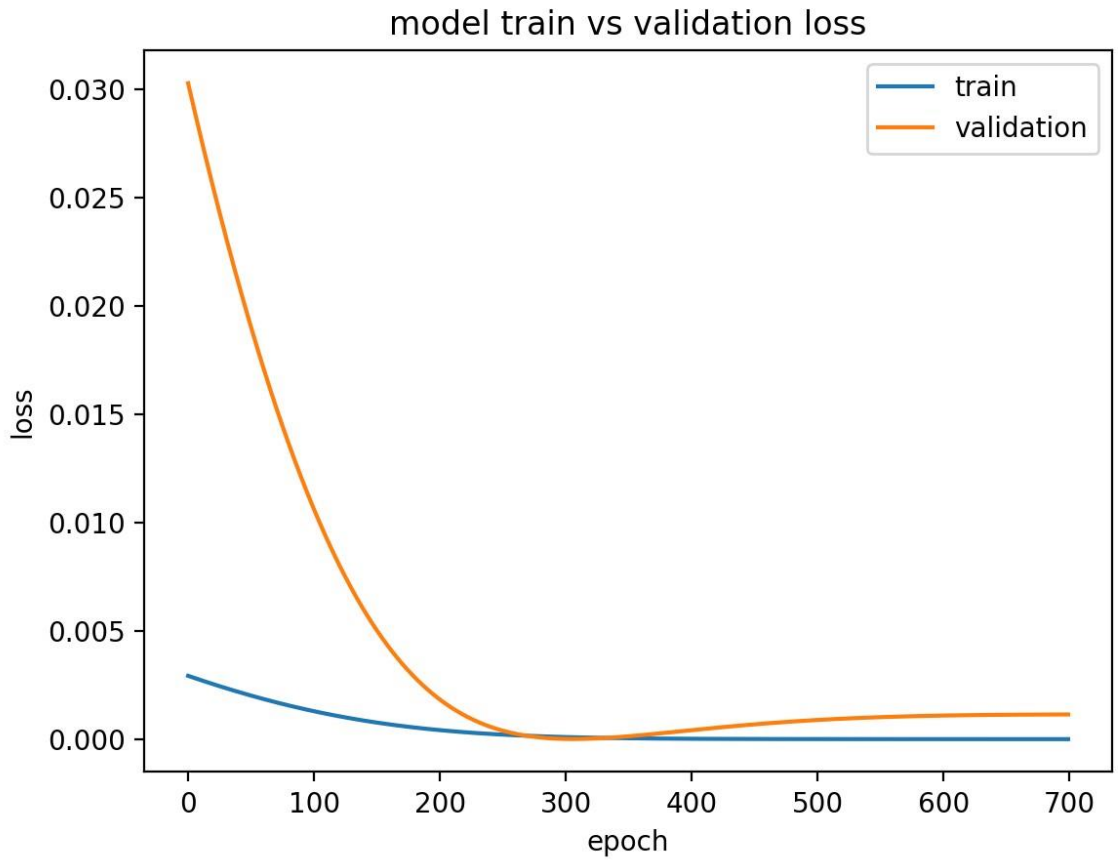


Figure 4.24: Representation of LSTM Over fitted Model Losses Graphically

4.4 Discussion

In this work the model performance are controlled and changed by different parameters, which are numbers of LSTM layer, number of the hidden layer, number of epochs, dropout layer, activation function and learning rate. So all the above parameters are adjusted and controlled until the models are Good fitted. If training model is over fitting, the network gets small errors during training and testing that does not manage to generalize to the dataset's patterns. Thus, when new inputs are given to the network, it cannot predict them accurately and gets significant deviations from the expected values. The model performance are also up down until we get the best accuracy by varying the above parameter to prevent the model from over fit and under fit.

The model used different Error optimization techniques to get a good fitted model and improve model performance. The above parameters are used to optimize the model performance, among the above parameter drop out some hidden layer output, so we had added drop out layer 0.04 at the first LSTM layer and added drop out 0.02 at second Layer that means 4% at layer one and 2% at layer two hidden layers output releases to increase the model performance and to reduce model error. Number of Epochs or number of training cycle are the main contribution for improving the model performance, so at epoch 10 to epoch 50 the model performance are very low and the result shows that the Loss graphs have been indicated that under fitted, the Epoch from 70 to 500 are the model performance are 96% and the error are 4%,so at this moment the loss graphic curve is indicated that the model are best fitted or a good fitted.

At the time of varying the drop out layer, we have gotten Under fitted and over fitted at any point except these the above points and also the error are increase to up to 40%.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Malaria is a world public health problem that causes massive morbidity and mortality that poses a higher burden of disease and affects all stages of the people. In Ethio-Amhara Region the malaria are widespread among all geographical zones . Usually, the majority of malaria diseases are created at the starting and end of summer, these months are comfortable for mosquito reproduction and for their Larva survival. The higher number of mosquito's results a higher number of malarial spreads in least developing country including in Ethio-Amhara Region.

To solve such kinds of problem ML are one of them. Malaria occurs all year round that proportional to weather and their month, these indicate a time series. LSTM are one types of model that solve time series problem recently that trained and predict future malaria using the inputted weather variable. So if we set the weather the model generates malaria according to the pre-trained data.

Malaria requires the use of methods such as fundamental, technical, and time series analysis in trying to predict the future disease. These methods generally point to a trend on future performance. Without a more accurate prediction method, the inputted weather data are unable to get the required malaria disease and may not be able to provide the requisite advice to their clients. There however exist Artificial Intelligence (AI) methods that can be used in developing prediction tools with the capability of indicating the most likely disease for the future.

The research are looked at the possibility of using AI to develop these predictive analyses.

The objectives of the research were to investigate AI models that were capable of malaria prediction, so as to develop something similar for malaria in the local context. The research also aimed at developing the LSTM based model for train and test the data. The final objective was to test the model and evaluate its performance. This was made possible by the use of actual malaria data that was compiled from authorized sources from the period 2006 to 2017.

If the people are to use these AI-based model, which have high accuracy, and next month and year Prediction capabilities, instead of trends, more people can be take care and also gain full knowledge about malaria transmission season to protect their selves from the disease. This increased the number of healthy people and malaria-free place for economic growth. The major contributions of this research are an ANN model that realizes high model performance and low model error in prediction, a model with comparatively low MSE when

measured against a similar model, a generic model that is applicable at any malaria predictive and a new baseline model that can be subjected to further tweaking by other researchers.

The AI method that was found suitable for developing such predictive models was the LSTM, which exploits parallel computing to gain intelligence from the input data as a basis of predicting future values. One challenged with the model-based systems is the ability to formulate an optimal network, with the right number of layers and the right number of neurons at each of the layers. Other considerations such as training cycles, activation function, stopping conditions, data volumes, and data partitioning, also need to the formulated and determined. This research used experimental evaluation to determine these parameters, to come up with a model that could suit the malaria prediction problem.

5.2 Recommendations for Further Research

There is a need to formulate a model that is generic enough to suit the full range of Malaria prediction at the Federal Level. This work selects W/Go jam and S/Gondar Zone in Amhara Region for study in addition to the whole Amhara Region and twelve years of data as the basis of the study. There is also a need to explore the effect of LSTM configurations by adding more hidden layer studied in this research and Finally, further research is needed to determine how long the trained LSTM system remains valid and effective in prediction additionally add more data from the federal public health institute and Ethiopian National Metrological Agency to Expand the prediction in the country level for more accurate training and prediction. Change the monthly data to daily for prediction to increase the dataset and the model performance.

REFERENCES

D. G. Ayele, T. T. Zewotir, and H. G. Mwambi, "Prevalence and risk factors of malaria in Ethiopia," pp. 1–9, 2012.

S. Thakur and R. Dharavath, "Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach," *Clin. Epidemiol. Glob. Heal.*, no. January, pp. 1–6, 2018.

A. Review, "APHI , PHEM Directorate Malaria Prevention control and Elimination Case Team," 2010.

E. Silva and J. Cesar, "Prediction of assets behavior in financial series using machine learning algorithms," vol. 2, no. 11, pp. 46–52, 2013.

T. Deress, "Plasmodium falciparum and Plasmodium vivax Prevalence in Ethiopia : A Systematic Review and Meta-Analysis," vol. 2019, 2019.

H. S. Taffese *et al.*, "Malaria epidemiology and interventions in Ethiopia from 2001 to 2016," pp. 1–9, 2018.

M. Sewe, J. Rocklöv, J. Williamson, M. Hamel, and A. Nyaguara, "The Association of Weather Variability and Under Five Malaria Mortality in KEMRI / CDC HDSS in Western Kenya 2003 to 2008 : A Time Series Analysis," pp. 1983–1997, 2015.

M. Taye, F. Zewdu, and D. Ayalew, "Characterizing the Climate System of Western Amhara , Ethiopia : A GIS Approach," vol. 1, no. 10, pp. 319–355, 2013.

E. Loha and T. M. Lunde, "Predicting Malaria Epidemics in Ethiopia," 2010.

"malaria pridection in nural network.pdf." .

A. Arifianto, "Malaria Incidence Forecasting from Incidence Record and Weather Pattern Using Polynomial Neural Network," *Int. J. Futur. Comput. Commun.*, pp. 60–65, 2014.

D. R. Tefera, S. O. Sinkie, and D. W. Daka, "Economic Burden of Malaria and Associated Factors Among Rural Households in Chewaka District , Western Ethiopia," 2020.

D. Lankir, S. Solomon, and A. Gize, "A five-year trend analysis of malaria surveillance data in selected zones of Amhara region , Northwest Ethiopia," pp. 1–9, 2020.

A. Alelign, Z. Tekeste, and B. Petros, "Prevalence of malaria in Woreta town , Amhara region , Northwest Ethiopia over eight years," pp. 1–6, 2018.

R. Parveen *et al.*, "Prediction of Malaria using Artificial Neural Network," *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 12, pp. 79–86, 2017.

E. Gofer, "Machine Learning Algorithms with Applications in Finance," no. March, 2014.

J. E. Gimnig, A. W. Hightower, and W. A. Hawley, "Application of geographic information systems to the study of the ecology of mosquitoes and mosquito-borne diseases," pp. 27–39.

"An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal."

A. Gupta, “Neural Networks in Data Processing,” vol. 5, no. 5, pp. 1638–1642, 2016. “2018 AI predictions,” pp. 1–25, 2018.

M. P. Darji, “Rainfall Forecasting Using Neural Network : A Survey,” pp. 706–713, 2015.

K. Y. Lee, N. Chung, and S. Hwang, “Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas,” *Ecol. Inform.*, vol. 36, pp. 172–180, 2016.

M. Obach, R. Wagner, H. Werner, and H. H. Schmidt, “Modelling population dynamics of aquatic insects with artificial neural networks,” *Ecol. Modell.*, vol. 146, no. 1–3, pp. 207–217, 2001.

G. Di Franco and M. Santurro, “research,” *Qual. Quant.*, no. 0123456789, 2020.

A. Dey and A. S. Learning, “Machine Learning Algorithms : A Review,” vol. 7, no. 3, pp. 1174–1179, 2016.

M. Selvam, B. Maniam, D. Kannaiah, C. Kathiravan, and T. Vadivel, “BIG DATA ANALYTICS-APPLICATION OF ARTIFICIAL NEURAL NETWORK IN FORECASTING,” vol. 22, no. 3, pp. 1–13, 2018.

S. N. Kavitha, C. T. S, and P. Jannu, “Supervised Machine Learning Approaches : A Survey,” vol. 5, no. 2, pp. 44–52, 2018.

P. Ezanno *et al.*, “A generic weather-driven model to predict mosquito population dynamics applied to species of Anopheles, Culex and Aedes genera of southern France,” *Prev. Vet. Med.*, vol. 120, no. 1, pp. 39–50, 2015.

V. Sharma, A. Kumar, L. Panat, and G. Karajkhede, “Malaria Outbreak Prediction Model Using Machine Learning,” vol. 4, no. 12, pp. 4415–4419, 2015.

I. Kiskin *et al.*, “Mosquito Detection with Neural Networks: The Buzz of Deep Learning,” pp. 1–16, 2017.

W. B. Abegaz and A. Mekoya, “Rainfall Variability and Trends over Central,” vol. 24, no. 4, 2020.

A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long ShortTerm Memory (LSTM) Network,” vol. 404, no. March, pp. 1–43, 2020.

H. Apaydin, H. Feizi, M. T. Sattari, and M. S. Colak, “Comparative Analysis of Recurrent Neural Network,” pp. 1–18.

“Introduction to Deep Learning.”

J. Brownlee, “Long Short-Term Memory Networks With Python Develop Sequence Prediction Models With Deep Learning.”

T. N. Sainath, O. Vinyals, A. Senior, and N. York, “No Title,” pp. 1–5.

Deep Learning_ Recurrent Neural Networks in Python_ LSTM, GRU, and more RNN machine learning architectures in Python and Theano (Machine Learning in Python) (PDFDrive).” .

M. Luong, “M - s s l,” no. c, pp. 1–10, 2016.

R. NeuralCascade-correlation and S. Chunking, “2 PREVIOUS WORK,” vol. 9, no. 8, pp. 1–32, 1997.

I. Sutskever, “Sequence to Sequence Learning with Neural Networks,” pp. 1–9.

Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks : LSTM Cells and Network Architectures,” vol. 1270, pp. 1235–1270, 2019.

R. C. Staudemeyer and E. R. Morris, “– Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks,” no. September, 2019.

W. Forecasting, M. For, and C. Datacenters, “Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based ScienceDirect Short Term Memory Recurrent Based Workload Forecasting Model For Cloud Datacenters” *Procedia Comput. Sci.*, vol. 125, no. January, pp. 676–682, 2018.

C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, “Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation,” pp. 1–16, 2018.

I. Islek, “Use of LSTM for Short-Term and Long-Term Travel Time Prediction.”

R. Staudemeyer and E. Morris, *Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks*. 2019.

D. Monner and J. A. Reggia, “A generalized LSTM-like training algorithm for secondorder recurrent neural networks,” *Neural Networks*, vol. 25, pp. 70–83, 2012.

A. Elsaid, B. Wild, J. Higgins, and T. Desell, “Using LSTM Recurrent Neural Networks to Predict Excess Vibration Events in Aircraft Engines,” pp. 260–269, 2016.

F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to Forget: Continual Prediction with LSTM,” *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

F. A. Gers and F. Cummins, “1 Introduction 2 Standard LSTM,” pp. 1–19, 1999.

F. A. Gers and N. N. Schraudolph, “Learning Precise Timing with LSTM Recurrent Networks,” vol. 3, pp. 115–143, 2002.

M. A. Mandale and B. A. Jadhawar, “Weather forecast prediction : a Data Mining application,” vol. 3, no. 2, pp. 1279–1284, 2015.

https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstmmodels/

APPENDIXES

Zones	Monthly											
Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	8952	9078	10980	10993	15897	16798	14968	16987	18967	20955	25968	32968
2007	4968	6987	7894	7968	8967	9855	4698	4970	6087	7098	7128	8058
2008	6047	6601	7085	7900	8021	8968	5023	5039	6069	6098	7098	9266
2009	6075	6987	7968	8936	9075	9687	6987	7098	7968	8967	9697	9978
2010	986	1024	1896	1926	2022	2637	1632	1680	2022	2967	3045	3601
2011	1560	1600	1670	1980	3047	4078	2069	2079	3047	4967	5697	5699
2012	2631	3696	3901	4396	4474	4896	1967	3697	4967	5967	6042	6098
2013	369	478	569	698	2074	3450	501	692	1690	2690	3697	4967
2014	2045	2967	4369	4967	5697	5960	3314	3496	3967	4967	5001	5221
2015	2961	3890	4967	5021	5967	6042	2361	2697	3697	4023	4367	4967
2016	360	569	697	1714	2097	3496	697	798	601	796	3697	4967
2017	1652	1896	2047	2690	2997	3967	1967	2049	3697	4967	4997	5096

Appendix 1: South Gondar Monthly Malaria Cassese (2006-2017)

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOv	DEC
2006	3025	2789	2047	2781	10235	19825	6789	4520	70828	98056	37890	67580
2007	2045	1907	4097	2098	9478	24896	7623	3785	20789	19080	8795	6078
2008	3902	5092	2096	5967	7341	14789	9710	6065	10046	13985	11963	8967
2009	2825	3940	5901	7900	9700	6781	4071	6077	10298	20949	30059	4905
2010	1091	2089	1245	3961	2701	4211	5068	2094	3752	1045	2701	1648
2011	950	1022	3097	2046	4980	6712	3041	2078	9740	8120	2070	4071
2012	3045	2987	3032	2098	4017	5714	4960	3021	3960	4071	6981	5320
2013	4039	5027	4090	5874	6790	7048	6897	4968	6998	7789	6786	5067
2014	1402	2078	2985	3014	3520	3690	1224	1822	2487	2698	4870	6980
2015	3910	4078	4099	5268	5870	6640	7820	7900	8011	8079	9780	20198
2016	2078	3984	4520	5690	6789	6790	4970	4980	5930	6923	7047	8967
2017	1087	1901	2968	3785	3988	4012	2631	3487	4396	4399	5098	5021

Appendix 2: West Go jam Monthly Malaria Cases (2006-2017)

			Reg; Gojamm Station; B.Dar Element Monthly Tmax									
MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	27.6	29.0	29.2	29.8	28.2	26.5	24.2	23.7	24.5	26.4	26.2	25.7
2007	25.8	27.9	30.0	29.7	29.5	25.8	23.6	24.2	24.9	26.3	26.5	27.0
2008	27.8	28.6	30.4	28.6	27.7	26.2	24.2	24.8	25.4	25.9	26.1	26.3
2009	26.8	28.9	30.1	30.5	30.6	28.6	14.9	24.4	26.2	26.3	27.3	26.4
2010	27.1	28.6	29.7	31.0	28.9	27.5	24.0	23.9	25.0	26.7	26.6	25.9
2011	26.3	29.3	29.2	30.8	28.7	26.9	24.8	24.4	25.1	26.5	26.0	25.9
2012	27.5	29.7	30.6	30.2	30.6	28.1	24.6	24.3	25.2	26.6	26.7	27.2
2013	27.9	30.2	31.1	31.8	30.6	35.2	25.7	24.7	26.9	26.9	27.4	26.9
2014	27.8	28.6	29.6	29.6	28.9	28.5	27.1	26.3	26.1	26.6	27.3	27.1
2015	27.0	30.5	31.0	31.4	29.9	28.4	27.1	26.8	27.3	28.1	28.0	27.3
2016	27.7	29.5	31.8	30.9	28.9	27.9	25.8	25.8	26.8	27.9	28.1	28.3
2017	28.7	28.3	30.0	30.8	28.3	28.1	26.0	25.5	26.8	27.7	27.2	26.8

Appendix3: West Go jam Monthly T Max(2006-2017)

			Station; B.Dar Element : Monthly Minimum temprature							**LAT.**	**11.590**
										ELEV.	**1800**
MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	8.7	10.9	12.2	14.1	15.6	15.0	15.0	14.8	14.1	14.3	11.6	8.6
2007	7.3	8.5	10.6	12.2	13.1	11.2	10.7	10.1	9.6	13.7	10.7	6.3
2008	9.0	8.5	9.1	15.1	15.0	17.9	13.3	13.2	12.5	12.5	8.7	7.6
2009	5.4	11.4	12.9	13.1	*	14.8	14.1	14.9	12.1	12.0	8.6	8.7
2010	8.4	9.9	12.4	16.7	16.0	14.4	14.7	14.6	14.0	13.4	9.8	5.9
2011	6.4	6.4	11.4	15.7	15.0	14.4	13.6	13.2	12.9	12.1	10.5	6.0
2012	8.0	8.9	12.5	12.1	15.5	15.4	14.4	13.9	13.6	12.8	12.8	9.2
2013	7.9	9.3	14.6	13.1	16.3	4.5	14.4	14.3	13.9	14.1	12.3	7.7
2014	8.7	8.0	12.8	14.4	14.4	13.3	13.2	13.1	14.3	15.0	12.6	9.4
2015	8.4	10.6	14.1	14.8	16.3	15.8	14.9	14.7	14.8	15.0	13.7	12.3
2016	8.5	10.3	13.2	14.1	16.0	14.9	14.0	14.4	14.1	14.2	8.8	7.6
2017	6.0	13.0	13.1	16.6	16.6	15.4	15.1	14.9	14.6	15.0	12.1	7.4

Appendix 4: West Go jam Monthly T Min (2006-2017)

Reg; Gojamm Station; B.Dar Element : Monthly RH												
MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	43	39	36	38	62	70	81	81	69	55	46	45
2007	53	57	47	48	52	67	79	79	66	64	58	49
2008	47	51	43	50	59	70	78		73	65	56	50
2009												
2010	34	44	43	40	43	56	67	79	72	70	61	59
								181				
	50	43	38	43	59	68	80	82	77	66	58	54
2011	55	34	45	40	57	70	76	81	77	63	60	51
2012	43	33	37	30	46	61	76	78	75	62	63	55
2013	46	37	36	34	53	65	77	82	74	70	58	47
2014	45	41	47	51	66	67	76	77	74	67	55	51
2015	49	39	38	35	59	68	73	75	74	65	61	65
2016	54	44	41	39	63	69	80	77	75	66	49	45
2017	37	53	42	47	64	67	76	78	73	69	58	76

Appendix 5: West Gojam Monthly RH (2006-2017)

Reg; Gojam
Station; B.Dar
Element : Monthly Rain fall

LONG.	**37.388**
LAT.	**11.590**
Elev	**1800**

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	3.1	0.2	0.1	6.7	151.2	225.5	563.9	364.1	211.0	153.7	0.0	3.7
2007	0.0	0.0	1.1	29.2	16.2	285.6	314.8	328.8	203.4	115.6	11.4	0.0
2008	1.8	0.0	0.0	104.3	87.8	175.6	481.5	337.6	150.2	56.5	33.1	0.0
2009	0.0	0.0	7.7	3.0	8.0	66.3	319.5	618.5	112.1	56.8	3.0	0.0
2010	13.3	0.0	0.0	34.0	72.1	127.3	407.8	449.3	182.2	54.6	1.5	0.0
2011	0.0	0.0	28.4	12.9	103.0	169.0	415.4	312.8	144.0	37.9	28.1	0.0
2012	0.0	0.0	1.0	0.0	25.4	122.0	466.5	504.4	255.9	7.6	2.0	11.2
2013	0.0	0.0	1.4	1.4	88.0	148.6	594.0	350.3	137.9	169.1	16.6	0.0
2014	0.0	0.0	65.9	66.6	163.7	178.4	378.4	480.8	260.0	117.4	0.0	0.4
2015	0.0	0.8	0.4	0.0	136.8	89.3	302.2	248.9	223.9	116.7	12.2	31.8
2016	0.0	0.0	23.8	8.5	171.2	248.8	409.6	259.9	274.4	104.8	0.5	0.0
2017	0.0	25.9	2.6	41.9	136.8	50.0	528.7	485.7	189.2	128.9	9.6	0.0

Appendix 6: West Go jam Monthly RR (2006-2017)

Region
Station Debretabor Alt.
Element_ Monthly Humidity

MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	31.4	33.1	33.1	42.9	61.1	70.4	82.0	71.0	77.9	48.4	40.6	43.3
2007	39.0	38.0	32.2	43.8	51.5	75.3	85.0	80.5	73.0	54.5	44.6	34.3
2008	45.5	34.2	20.8	48.4	69.3	71.8	81.2	79.2	71.3	62.2	49.6	40.4
2009	42.6	49.1	27.7	39.2	0.0	59.4	77.5	225.1	67.5	66.0	50.3	49.8
2010	44.2	39.1	40.3	48.7	54.2	70.9	80.4	83.1	75.2	59.9	52.1	49.4
2011	*	30.4	41.0	37.4	58.7	67.0	77.8	81.8	78.7	55.2	57.2	50.0
2012	41.7	29.4	40.2	35.5	46.3	68.0	82.7	79.5	76.6	54.9	64.1	49.2
2013	40.9	33.7	34.8	41.5	57.5	na	76.6	83.5	72.3	76.6	58.7	45.7
2014	43.6	50.5	48.2	49.7	62.9	61.5	72.7	77.3	73.0	67.5	58.7	56.0
2015	42.6	41.2	42.0	37.8	63.3	71.3	78.0	80.9	75.1	62.2	61.2	66.2
2016	48.9	42.7	35.6	35.2	69.0	69.1	78.9	81.2	79.0	66.2	63.1	44.5
2017	29.0	45.6	38.1	48.4	68.1	68.5	80.7	85.3	74.1	68.0	54.6	39.9

Appendix 7: South Gondar Monthly Humidity (2006-2017)

Region Gondar
Station Debretabor Alt.
Monthly Rainfall

MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	0.0	1.4	6.8	63.2	147.3	170.0	482.2	452.5	255.0	47.5	0.0	7.9
2007	19.9	0.5	22.2	87.8	65.6	281.4	424.7	439.1	183.1	8.1	0.0	0.0
2008	81.4	0.0	1.5	81.9	211.5	209.4	376.4	341.8	228.6	51.8	2.5	18.5
2009	0.0	5.1	63.2	19.1	28.2	66.8	418.3	667.4	113.2	107.4	3.0	2.0
2010	13.1	0.0	33.3	52.1	65.3	151.2	499.3	527.9	203.0	41.4	21.1	9.7
2011	0.0	0.0	43.9	20.9	175.9	132.9	359.6	392.2	259.7	50.4	86.6	0.0
2012	0.0	0.0	33.8	0.0	57.2	277.7	389.3	447.7	214.0	24.4	41.3	4.0
2013	2.1	4.2	26.9	34.3	165.0	*	423.0	439.1	191.3	176.4	33.9	5.5
2014	5.4	4.3	151.5	63.7	206.3	165.2	340.8	453.6	222.2	86.1	50.8	0.0
2015	0	4.4	17.9	8.3	176.3	129.2	234.1	284.2	200.5	26.6	83.5	37.6
2016	0.0	0.0	16.6	16.6	193.0	162.3	375.6	398.8	168.4	27.9	1.5	0.0
2017	0	60	26.9	86.5	176.4	84.4	346	291	152	56.4	21.3	0

Appendix 8: South Gondar Monthly RR (2006-2017)

Region Gondar

Station Debretabor Alt.

Element Monthly Maximu temprature

		D/Tabor T max										
MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	23.7	25.1	25.2	24.4	22.6	21.6	19.5	19.3	20.7	22.8	22.4	6.4
2007	22.6	24.5	26.0	24.9	24.7	20.9	18.1	24.2	20.5	21.5	22.5	22.6
2008	22.8	23.6	25.6	24.2	22.2	20.6	19.2	19.6	20.5	21.7	22.0	21.8
2009	22.9	24.1	24.8	25.0	25.4	24.2	19.7	18.3	21.6	21.4	23.1	22.5
2010	22.5	24.4	24.7	24.7	23.8	21.6	18.9	19.2	20.5	21.9	22.2	21.6
2011	0.0	24.7	23.3	25.1	23.2	21.9	19.8	19.3	20.3	22.3	21.5	21.9
2012	23.1	25.0	25.7	25.8	25.6	23.4	21.2	20.1	21.1	23.4	23.1	24.1
2013	25.3	26.0	25.7	26.6	9.0	*	20.4	19.2	21.3	21.5	22.5	22.6
2014	23.3	25.2	24.7	25.4	24.2	23.9	21.9	20.0	21.1	21.3	22.1	22.0
2015	22.9	25.3	25.7	26.5	24.1	22.4	20.8	20.2	20.9	22.8	22.2	21.2
2016	22.5	24.9	26.0	25.9	22.6	22.2	19.3	19.5	20.5	20.8	22.6	22.7
2017	23.7	24.4	24.6	25.1	22.4	23.5	20.6	18.7	21.3	21.7	22.0	22.2

Appendix 9: South Gondar Monthly T max (2006-2017)

Region Gondar

Station Debretabor Alt.

Element Monthly Minimum temprature

MONTH	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	8.5	9.9	10.7	10.5	10.4	10.3	10.4	10.3	9.5	9.3	8.0	2.4
2007	8.2	9.5	10.7	11.1	11.6	10.9	10.5	10.1	9.7	7.9	7.5	7.4
2008	8.4	8.9	10.0	10.7	10.9	10.0	10.0	10.5	9.5	9.1	8.0	8.0
2009	8.0	10.1	10.6	11.5	11.6	10.8	10.6	11.6	9.6	8.6	8.3	8.0
2010	8.3	9.6	8.7	6.7	5.2	11.8	10.0	9.3	9.3	9.0	8.4	8.0
2011	0.0	9.0	10.1	9.6	8.7	8.8	8.4	9.7	9.1	7.8	8.3	7.5
2012	7.9	9.1	10.4	11.2	11.6	10.1	9.8	9.7	9.2	7.9	8.1	6.9
2013	8.0	9.5	10.1	9.5	11.5	*	9.9	9.7	9.4	8.9	8.4	6.9
2014	8.2	8.9	10.5	10.5	10.8	10.2	10.6	10.2	9.5	9.3	8.3	7.9
2015	7.6	10.0	11.0	11.9	11.5	11.1	10.5	10.5	9.2	8.8	9.0	8.6
2016	7.2	9.1	12.5	12.5	11.4	10.9	9.6	10.3	9.6	9.2	7.8	7.6
2017	6.6	8.6	11.2	11.5	10.9	11.4	10.8	11.0	10.6	9.9	8.1	7.2

Appendix 10: South Gondar Monthly T min (2006-2017)

year	month	T max	RR	humidity	Tmin	Total Malaria cases
2006	JAN	23.7	0.0	31.4	8.5	8952
2006	FEB	25.1	1.4	33.1	9.9	9078
2006	MAR	25.2	6.8	33.1	10.7	10980
2006	APR	24.4	63.2	42.9	10.5	10993
2006	MAY	22.6	147.3	61.1	10.4	15897
2006	JUN	21.6	170.0	70.4	10.3	16798
2006	JUL	19.5	482.2	82.0	10.4	14968
2006	AUG	19.3	452.5	71.0	10.3	16987
2006	SEP	20.7	255.0	77.9	9.5	18967
2006	OCT	22.8	47.5	48.4	9.3	20955
2006	NOV	22.4	0.0	40.6	8.0	25968
2006	DEC	6.4	7.9	43.3	2.4	32968
2007	JAN	22.6	19.9	39.0	8.2	4968
2007	FEB	24.5	0.5	38.0	9.5	6987
2007	MAR	26.0	22.2	32.2	10.7	7894
2007	APR	24.9	87.8	43.8	11.1	7968
2007	MAY	24.7	65.6	51.5	11.6	8967
2007	JUN	20.9	281.4	75.3	10.9	9855
2007	JUL	18.1	424.7	85.0	10.5	4698
2007	AUG	24.2	439.1	80.5	10.1	4970
2007	SEP	20.5	183.1	73.0	9.7	6087
2007	OCT	21.5	8.1	54.5	7.9	7098
2007	NOV	22.5	0.0	44.6	7.5	7128
2007	DEC	22.6	0.0	34.3	7.4	8058
2008	JAN	22.8	81.4	45.5	8.4	6047
2008	FEB	23.6	0.0	34.2	8.9	6601
2008	MAR	25.6	1.5	20.8	10.0	7085
2008	APR	24.2	81.9	48.4	10.7	7900
2008	MAY	22.2	211.5	69.3	10.9	8021
2008	JUN	20.6	209.4	71.8	10.0	8968
2008	JUL	19.2	376.4	81.2	10.0	5023
2008	AUG	19.6	341.8	79.2	10.5	5039
2008	SEP	20.5	228.6	71.3	9.5	6069
2008	OCT	21.7	51.8	62.2	9.1	6098
2008	NOV	22.0	2.5	49.6	8.0	7098
2008	DEC	21.8	18.5	40.4	8.0	9266
2009	JAN	22.9	0.0	42.6	8.0	6075
2009	FEB	24.1	5.1	49.1	10.1	6987

2009	MAR	24.8	63.2	27.7	10.6	7968
2009	APR	25.0	19.1	39.2	11.5	8936
2009	MAY	25.4	28.2	0.0	11.6	9075
2009	JUN	24.2	66.8	59.4	10.8	9687
2009	JUL	19.7	418.3	77.5	10.6	6987
2009	AUG	18.3	667.4	225.1	11.6	7098
2009	SEP	21.6	113.2	67.5	9.6	7968
2009	OCT	21.4	107.4	66.0	8.6	8967
2009	NOV	23.1	3.0	50.3	8.3	9697
2009	DEC	22.5	2.0	49.8	8.0	9978
2010	JAN	22.5	13.1	44.2	8.3	986
2010	FEB	24.4	0.0	39.1	9.6	1024
2010	MAR	24.7	33.3	40.3	8.7	1896
2010	APR	24.7	52.1	48.7	6.7	1926
2010	MAY	23.8	65.3	54.2	5.2	2022
2010	JUN	21.6	151.2	70.9	11.8	2637
2010	JUL	18.9	499.3	80.4	10.0	1632
2010	AUG	19.2	527.9	83.1	9.3	1680
2010	SEP	20.5	203.0	75.2	9.3	2022
2010	OCT	21.9	41.4	59.9	9.0	2967
2010	NOV	22.2	21.1	52.1	8.4	3045
2010	DEC	21.6	9.7	49.4	8.0	3601
2012	JAN	0.0	0.0	40.0	0.0	1560
2012	FEB	24.7	0.0	30.4	9.0	1600
2012	MAR	23.3	43.9	41.0	10.1	1670
2012	APR	25.1	20.9	37.4	9.6	1980
2012	MAY	23.2	175.9	58.7	8.7	3047
2012	JUN	21.9	132.9	67.0	8.8	4078
2012	JUL	19.8	359.6	77.8	8.4	2069
2012	AUG	19.3	392.2	81.8	9.7	2079
2012	SEP	20.3	259.7	78.7	9.1	3047
2012	OCT	22.3	50.4	55.2	7.8	4967
2012	NOV	21.5	86.6	57.2	8.3	5697
2012	DEC	21.9	0.0	50.0	7.5	5699
2013	JAN	23.1	0.0	41.7	7.9	2631
2013	FEB	25.0	0.0	29.4	9.1	3696
2013	MAR	25.7	33.8	40.2	10.4	3901
2013	APR	25.8	0.0	35.5	11.2	4396
2013	MAY	25.6	57.2	46.3	11.6	4474
2013	JUN	23.4	277.7	68.0	10.1	4896
2013	JUL	21.2	389.3	82.7	9.8	1967

2013	AUG	20.1	447.7	79.5	9.7	3697
2013	SEP	21.1	214.0	76.6	9.2	4967
2013	OCT	23.4	24.4	54.9	7.9	5967
2013	NOV	23.1	41.3	64.1	8.1	6042
2013	DEC	24.1	4.0	49.2	6.9	6098
2013	JAN	25.3	2.1	40.9	8.0	369
2013	FEB	26.0	4.2	33.7	9.5	478
2013	MAR	25.7	26.9	34.8	10.1	569
2013	APR	26.6	34.3	41.5	9.5	698
2013	MAY	9.0	165.0	57.5	11.5	2074
2013	JUN	14.8	290.0	na	10.0	3450
2013	JUL	20.4	423.0	76.6	9.9	501
2013	AUG	19.2	439.1	83.5	9.7	692
2013	SEP	21.3	191.3	72.3	9.4	1690
2013	OCT	21.5	176.4	76.6	8.9	2690
2013	NOV	22.5	33.9	58.7	8.4	3697
2013	DEC	22.6	5.5	45.7	6.9	4967
2014	JAN	23.3	5.4	43.6	8.2	2045
2014	FEB	25.2	4.3	50.5	8.9	2967
2014	MAR	24.7	151.5	48.2	10.5	4369
2014	APR	25.4	63.7	49.7	10.5	4967
2014	MAY	24.2	206.3	62.9	10.8	5697
2014	JUN	23.9	165.2	61.5	10.2	5960
2014	JUL	21.9	340.8	72.7	10.6	3314
2014	AUG	20.0	453.6	77.3	10.2	3496
2014	SEP	21.1	222.2	73.0	9.5	3967
2014	OCT	21.3	86.1	67.5	9.3	4967
2014	NOV	22.1	50.8	58.7	8.3	5001
2014	DEC	22.0	0.0	56.0	7.9	5221
2015	JAN	22.9	0	42.6	7.6	2961
2015	FEB	25.3	4.4	41.2	10.0	3890
2015	MAR	25.7	17.9	42.0	11.0	4967
2015	APR	26.5	8.3	37.8	11.9	5021
2015	MAY	24.1	176.3	63.3	11.5	5967
2015	JUN	22.4	129.2	71.3	11.1	6042
2015	JUL	20.8	234.1	78.0	10.5	2361
2015	AUG	20.2	284.2	80.9	10.5	2697
2015	SEP	20.9	200.5	75.1	9.2	3697
2015	OCT	22.8	26.6	62.2	8.8	4023
2015	NOV	22.2	83.5	61.2	9.0	4367
2015	DEC	21.2	37.6	66.2	8.6	4967

2016	JAN	22.5	0.0	48.9	7.2	360
2016	FEB	24.9	0.0	42.7	9.1	569
2016	MAR	26.0	16.6	35.6	12.5	697
2016	APR	25.9	16.6	35.2	12.5	1714
2016	MAY	22.6	193.0	69.0	11.4	2097
2016	JUN	22.2	162.3	69.1	10.9	3496
2016	JUL	19.3	375.6	78.9	9.6	697
2016	AUG	19.5	398.8	81.2	10.3	798
2016	SEP	20.5	168.4	79.0	9.6	601
2016	OCT	20.8	27.9	66.2	9.2	796
2016	NOV	22.6	1.5	63.1	7.8	3697
2016	DEC	22.7	0.0	44.5	7.6	4967
2017	JAN	23.7	0	29.0	6.6	1652
2017	FEB	24.4	60	45.6	8.6	1896
2017	MAR	24.6	26.9	38.1	11.2	2047
2017	APR	25.1	86.5	48.4	11.5	2690
2017	MAY	22.4	176.4	68.1	10.9	2997
2017	JUN	23.5	84.4	68.5	11.4	3967
2017	JUL	20.6	346	80.7	10.8	1967
2017	AUG	18.7	291	85.3	11.0	2049
2017	SEP	21.3	152	74.1	10.6	3697
2017	OCT	21.7	56.4	68.0	9.9	4967
2017	NOV	22.0	21.3	54.6	8.1	4997
2017	DEC	22.2	0	39.9	7.2	5096

Appendix 11: South Gondar Dataset all Attribute in one (2006-2017)

year	month	Tmax	RR	Humidity	Tmin	Monthly Malaria
2006	JAN	27.6	3.1	43	8.7	3025
2006	FEB	29.0	0.2	39	10.9	2789
2006	MAR	29.2	0.1	36	12.2	2047
2006	APR	29.8	6.7	38	14.1	2781
2006	MAY	28.2	151.2	62	15.6	10235
2006	JUN	26.5	225.5	70	15.0	19825
2006	JUL	24.2	563.9	81	15.0	6789
2006	AUG	23.7	364.1	81	14.8	4520
2006	SEP	24.5	211.0	69	14.1	70828
2006	OCT	26.4	153.7	55	14.3	98056
2006	NOV	26.2	0.0	46	11.6	37890

2006	DEC	25.7	3.7	45	8.6	67580
2007	JAN	25.8	0.0	53	7.3	2045
2007	FEB	27.9	0.0	57	8.5	1907
2007	MAR	30.0	1.1	47	10.6	4097
2007	APR	29.7	29.2	48	12.2	2098
2007	MAY	29.5	16.2	52	13.1	9478
2007	JUN	25.8	285.6	67	11.2	24896
2007	JUL	23.6	314.8	79	10.7	7623
2007	AUG	24.2	328.8	79	10.1	3785
2007	SEP	24.9	203.4	66	9.6	20789
2007	OCT	26.3	115.6	64	13.7	19080
2007	NOV	26.5	11.4	58	10.7	8795
2007	DEC	27.0	0.0	49	6.3	6078
2008	JAN	27.8	1.8	47	9.0	3902
2008	FEB	28.6	0.0	51	8.5	5092
2008	MAR	30.4	0.0	43	9.1	2096
2008	APR	28.6	104.3	50	15.1	5967
2008	MAY	27.7	87.8	59	15.0	7341
2008	JUN	26.2	175.6	70	17.9	14789
2008	JUL	24.2	481.5	78	13.3	9710
2008	AUG	24.8	337.6	79	13.2	6065
2008	SEP	25.4	150.2	73	12.5	10046
2008	OCT	25.9	56.5	65	12.5	13985
2008	NOV	26.1	33.1	56	8.7	11963
2008	DEC	26.3	0.0	50	7.6	8967
2009	JAN	26.8	0.0	34	5.4	2825
2009	FEB	28.9	0.0	44	11.4	3940
2009	MAR	30.1	7.7	43	12.9	5901
2009	APR	30.5	3.0	40	13.1	7900
2009	MAY	30.6	8.0	43	14.0	9700
2009	JUN	28.6	66.3	56	14.8	6781
2009	JUL	14.9	319.5	67	14.1	4071
2009	AUG	24.4	618.5	181	14.9	6077
2009	SEP	26.2	112.1	72	12.1	10298
2009	OCT	26.3	56.8	70	12.0	20949
2009	NOV	27.3	3.0	61	8.6	30059
2009	DEC	26.4	0.0	59	8.7	4905
2010	JAN	27.1	13.3	50	8.4	1091
2010	FEB	28.6	0.0	43	9.9	2089
2010	MAR	29.7	0.0	38	12.4	1245

2010	APR	31.0	34.0	43	16.7	3961
2010	MAY	28.9	72.1	59	16.0	2701
2010	JUN	27.5	127.3	68	14.4	4211
2010	JUL	24.0	407.8	80	14.7	5068
2010	AUG	23.9	449.3	82	14.6	2094
2010	SEP	25.0	182.2	77	14.0	3752
2010	OCT	26.7	54.6	66	13.4	1045
2010	NOV	26.6	1.5	58	9.8	2701
2010	DEC	25.9	0.0	54	5.9	1648
2011	JAN	26.3	13.3	55	6.4	950
2011	FEB	29.3	0.0	34	6.4	1022
2011	MAR	29.2	0.0	45	11.4	3097
2011	APR	30.8	34.0	40	15.7	2046
2011	MAY	28.7	72.1	57	15.0	4980
2011	JUN	26.9	127.3	70	14.4	6712
2011	JUL	24.8	407.8	76	13.6	3041
2011	AUG	24.4	449.3	81	13.2	2078
2011	SEP	25.1	182.2	77	12.9	9740
2011	OCT	26.5	54.6	63	12.1	8120
2011	NOV	26.0	1.5	60	10.5	2070
2011	DEC	25.9	0.0	51	6.0	4071
2012	JAN	27.5	0.0	43	8.0	3045
2012	FEB	29.7	0.0	33	8.9	2987
2012	MAR	30.6	28.4	37	12.5	3032
2012	APR	30.2	12.9	30	12.1	2098
2012	MAY	30.6	103.0	46	15.5	4017
2012	JUN	28.1	169.0	61	15.4	5714
2012	JUL	24.6	415.4	76	14.4	4960
2012	AUG	24.3	312.8	78	13.9	3021
2012	SEP	25.2	144.0	75	13.6	3960
2012	OCT	26.6	37.9	62	12.8	4071
2012	NOV	26.7	28.1	63	12.8	6981
2012	DEC	27.2	0.0	55	9.2	5320
2013	JAN	27.9	0.0	46	7.9	4039
2013	FEB	30.2	0.0	37	9.3	5027
2013	MAR	31.1	1.4	36	14.6	4090
2013	APR	31.8	1.4	34	13.1	5874
2013	MAY	30.6	88.0	53	16.3	6790
2013	JUN	35.2	148.6	65	4.5	7048
2013	JUL	25.7	594.0	77	14.4	6897

2013	AUG	24.7	350.3	82	14.3	4968
2013	SEP	26.9	137.9	74	13.9	6998
2013	OCT	26.9	169.1	70	14.1	7789
2013	NOV	27.4	16.6	58	12.3	6786
2013	DEC	26.9	0.0	47	7.7	5067
2014	JAN	27.8	0.0	45	8.7	1402
2014	FEB	28.6	0.0	41	8.0	2078
2014	MAR	29.6	65.9	47	12.8	2985
2014	APR	29.6	66.6	51	14.4	3014
2014	MAY	28.9	163.7	66	14.4	3520
2014	JUN	28.5	178.4	67	13.3	3690
2014	JUL	27.1	378.4	76	13.2	1224
2014	AUG	26.3	480.8	77	13.1	1822
2014	SEP	26.1	260.0	74	14.3	2487
2014	OCT	26.6	117.4	67	15.0	2698
2014	NOV	27.3	0.0	55	12.6	4870
2014	DEC	27.1	0.4	51	9.4	6980
2015	JAN	27.0	0.0	49	8.4	3910
2015	FEB	30.5	0.8	39	10.6	4078
2015	MAR	31.0	0.4	38	14.1	4099
2015	APR	31.4	0.0	35	14.8	5268
2015	MAY	29.9	136.8	59	16.3	5870
2015	JUN	28.4	89.3	68	15.8	6640
2015	JUL	27.1	302.2	73	14.9	7820
2015	AUG	26.8	248.9	75	14.7	7900
2015	SEP	27.3	223.9	74	14.8	8011
2015	OCT	28.1	116.7	65	15.0	8079
2015	NOV	28.0	12.2	61	13.7	9780
2015	DEC	27.3	31.8	65	12.3	20198
2016	JAN	27.7	0.0	54	8.5	2078
2016	FEB	29.5	0.0	44	10.3	3984
2016	MAR	31.8	23.8	41	13.2	4520
2016	APR	30.9	8.5	39	14.1	5690
2016	MAY	28.9	171.2	63	16.0	6789
2016	JUN	27.9	248.8	69	14.9	6790
2016	JUL	25.8	409.6	80	14.0	4970
2016	AUG	25.8	259.9	77	14.4	4980
2016	SEP	26.8	274.4	75	14.1	5930
2016	OCT	27.9	104.8	66	14.2	6923
2016	NOV	28.1	0.5	49	8.8	7047

2016	DEC	28.3	0.0	45	7.6	8967
2017	JAN	28.7	0.0	37	6.0	1087
2017	FEB	28.3	25.9	53	13.0	1901
2017	MAR	30.0	2.6	42	13.1	2968
2017	APR	30.8	41.9	47	16.6	3785
2017	MAY	28.3	136.8	64	16.6	3988
2017	JUN	28.1	50.0	67	15.4	4012
2017	JUL	26.0	528.7	76	15.1	2631
2017	AUG	25.5	485.7	78	14.9	3487
2017	SEP	26.8	189.2	73	14.6	4396
2017	OCT	27.7	128.9	69	15.0	4399
2017	NOV	27.2	9.6	58	12.1	5098
2017	DEC	26.8	0.0	76	7.4	5021

Appendix 12: West Go jam Dataset all Attribute in one (2006-2017)

year	month	Tmax	RR	Humidity	Tmin	Malaria
2006	JAN	0.891874	0.002098	0.146954	0.790203	0.263493621
2006	FEB	0.941676	0.010189	0.147198	0.860615	0.267357704
2006	MAR	0.945927	0.094696	0.190612	0.842005	0.325686948
2006	APR	0.916594	0.220707	0.27132	0.829842	0.326085623
2006	MAY	0.849577	0.25472	0.312651	0.822765	0.476478165
2006	JUN	0.810394	0.722505	0.364483	0.829842	0.504109421
2006	JUL	0.73323	0.678004	0.315322	0.823894	0.447988224
2006	AUG	0.724019	0.38208	0.345975	0.764244	0.509905545
2006	SEP	0.77583	0.071172	0.215135	0.743729	0.57062684
2006	OCT	0.855031	0	0.180541	0.640255	0.631593474
2006	NOV	0.840075	0.011837	0.192346	0.195242	0.785328754
2006	DEC	0.24142	0.029817	0.17314	0.653995	1
2007	JAN	0.849092	0.000749	0.168999	0.762145	0.141315015
2007	FEB	0.921013	0.033263	0.143041	0.857254	0.203232336
2007	MAR	0.978286	0.131555	0.194759	0.886898	0.231047596
2007	APR	0.934126	0.098292	0.228752	0.930696	0.233316977
2007	MAY	0.926536	0.421636	0.334571	0.870063	0.263953631

2007	JUN	0.785974	0.63635	0.377526	0.840703	0.291186212
2007	JUL	0.680146	0.657926	0.357604	0.808637	0.133034838
2007	AUG	0.908841	0.274348	0.324351	0.774933	0.141376349
2007	SEP	0.77082	0.012137	0.242224	0.635635	0.175631747
2007	OCT	0.808613	0	0.198017	0.60231	0.206636408
2007	NOV	0.844834	0	0.152447	0.590639	0.207556428
2007	DEC	0.848607	0.121966	0.202281	0.672097	0.236077036
2008	JAN	0.855152	0	0.151769	0.717343	0.174405054
2008	FEB	0.886015	0.002248	0.092418	0.803465	0.19139475
2008	MAR	0.962409	0.122715	0.215049	0.858773	0.206237733
2008	APR	0.910457	0.316901	0.30769	0.873002	0.2312316
2008	MAY	0.834549	0.313755	0.319168	0.803258	0.234942345
2008	JUN	0.775204	0.56398	0.360613	0.805017	0.263984298
2008	JUL	0.72111	0.512137	0.352014	0.838376	0.143001717
2008	AUG	0.734684	0.342523	0.316798	0.761572	0.143492395
2008	SEP	0.769192	0.077615	0.276337	0.726834	0.175079735
2008	OCT	0.815521	0.003746	0.220233	0.637316	0.175969087
2008	NOV	0.825673	0.02772	0.17959	0.644427	0.206636408
2008	DEC	0.820611	0	0.189193	0.645203	0.27312316
2009	JAN	0.858788	0.007642	0.218191	0.812535	0.175263739
2009	FEB	0.903838	0.094696	0.122975	0.84691	0.203232336
2009	MAR	0.933565	0.028619	0.174172	0.925644	0.233316977
2009	APR	0.940639	0.042254	0	0.930855	0.263002944
2009	MAY	0.955138	0.10009	0.263776	0.866589	0.267265702
2009	JUN	0.909831	0.626761	0.344561	0.847686	0.286034102
2009	JUL	0.739774	1	1	0.93173	0.203232336
2009	AUG	0.68657	0.169613	0.299914	0.767451	0.206636408
2009	SEP	0.809894	0.160923	0.293249	0.691234	0.233316977
2009	OCT	0.804613	0.004495	0.223343	0.662701	0.263953631
2009	NOV	0.866875	0.002997	0.221299	0.644686	0.286340775
2009	DEC	0.847153	0.019628	0.196359	0.661753	0.294958292
2010	JAN	0.845214	0	0.17376	0.765581	0.019197743
2010	FEB	0.916585	0.049895	0.178873	0.694337	0.020363101
2010	MAR	0.927748	0.078064	0.216382	0.54085	0.047105005
2010	APR	0.928115	0.097842	0.240791	0.419705	0.048025025

2010	MAY	0.894783	0.226551	0.314873	0.949427	0.050969087
2010	JUN	0.811522	0.748127	0.357317	0.799069	0.06982949
2010	JUL	0.711172	0.79098	0.369356	0.742199	0.039008832
2010	AUG	0.719535	0.304165	0.334275	0.745005	0.040480864
2010	SEP	0.769317	0.062032	0.266017	0.722783	0.050969087
2010	OCT	0.822308	0.031615	0.231489	0.669382	0.079949706
2010	NOV	0.832686	0.014534	0.219292	0.6421	0.082341757
2010	DEC	0.812612	0	0.177727	0	0.099392787
2011	JAN	0	0	0.134895	0.718913	0.036800785
2011	FEB	0.92799	0.065778	0.182313	0.809413	0.038027478
2011	MAR	0.877088	0.031316	0.166175	0.771727	0.04017419
2011	APR	0.941265	0.26356	0.260714	0.699767	0.04968106
2011	MAY	0.872604	0.199131	0.297692	0.707327	0.082403091
2011	JUN	0.822668	0.538807	0.345564	0.670028	0.114021099
2011	JUL	0.742198	0.587654	0.363623	0.773985	0.052410451
2011	AUG	0.726322	0.389122	0.349678	0.731109	0.052717125
2011	SEP	0.760927	0.075517	0.245091	0.622446	0.082403091
2011	OCT	0.836003	0.129757	0.254297	0.665908	0.141284347
2011	NOV	0.806011	0	0.222015	0.602017	0.163671492
2011	DEC	0.823883	0	0.18518	0.635893	0.163732826
2012	JAN	0.869574	0	0.130703	0.726742	0.069645486
2012	FEB	0.940427	0.050644	0.178581	0.831911	0.102306183
2012	MAR	0.96447	0	0.157584	0.90106	0.108592983
2012	APR	0.97082	0.085706	0.205676	0.932247	0.123773307
2012	MAY	0.960955	0.416092	0.301987	0.806465	0.126165358
2012	JUN	0.880276	0.583308	0.367637	0.789242	0.139106968
2012	JUL	0.796251	0.670812	0.353017	0.779416	0.049282385
2012	AUG	0.75359	0.320647	0.340495	0.740462	0.10233685
2012	SEP	0.793863	0.03656	0.244088	0.634083	0.141284347
2012	OCT	0.878647	0.061882	0.284955	0.649341	0.171951668
2012	NOV	0.866625	0.005993	0.218719	0.552108	0.174251717
2012	DEC	0.90569	0.003147	0.181883	0.638385	0.175969087
2013	JAN	0.950282	0.006293	0.14964	0.765008	0.000276006
2013	FEB	0.977771	0.040306	0.154651	0.81174	0.003618744
2013	MAR	0.966045	0.051393	0.184243	0.763354	0.00640947

2013	APR	1	0.247228	0.255554	0.921386	0.010365554
2013	MAY	0.336558	0.434522	#VALUE!	0.801655	0.052563788
2013	JUN	0.556043	0.633803	0.340404	0.79338	0.094762022
2013	JUL	0.768134	0.657926	0.371076	0.779933	0.004324092
2013	AUG	0.722564	0.286635	0.321093	0.755693	0.010181551
2013	SEP	0.799249	0.264309	0.340548	0.716576	0.040787537
2013	OCT	0.807559	0.050794	0.260962	0.672321	0.071454858
2013	NOV	0.846086	0.008241	0.203096	0.551849	0.10233685
2013	DEC	0.85091	0.008091	0.19378	0.657099	0.141284347
2014	JAN	0.875876	0.006443	0.224539	0.713473	0.051674436
2014	FEB	0.945662	0.227	0.214132	0.838376	0.079949706
2014	MAR	0.927142	0.095445	0.220677	0.841471	0.122945289
2014	APR	0.953287	0.30911	0.27949	0.869408	0.141284347
2014	MAY	0.91078	0.247528	0.273255	0.814749	0.163671492
2014	JUN	0.899311	0.510638	0.323061	0.849496	0.171736997
2014	JUL	0.82352	0.679652	0.343557	0.814844	0.090591266
2014	AUG	0.752621	0.332934	0.3245	0.759167	0.096172718
2014	SEP	0.792611	0.129008	0.299699	0.746056	0.110617026
2014	OCT	0.798554	0.076116	0.260666	0.663236	0.141284347
2014	NOV	0.830056	0	0.248674	0.629428	0.142327036
2014	DEC	0.827156	0	0.189107	0.60693	0.149073847
2015	JAN	0.858545	0.006593	0.183154	0.799651	0.079765702
2015	FEB	0.94919	0.02682	0.186814	0.8857	0.108255643
2015	MAR	0.966651	0.012436	0.168011	0.949961	0.141284347
2015	APR	0.995617	0.264159	0.281439	0.923713	0.142940383
2015	MAY	0.904115	0.193587	0.316887	0.890639	0.171951668
2015	JUN	0.843081	0.350764	0.346539	0.839152	0.174251717
2015	JUL	0.781223	0.425832	0.359266	0.840962	0.061365309
2015	AUG	0.757347	0.30042	0.333504	0.736187	0.071669529
2015	SEP	0.78422	0.039856	0.276308	0.708818	0.10233685
2015	OCT	0.857697	0.125112	0.271833	0.722826	0.112334396
2015	NOV	0.834941	0.056338	0.294138	0.689165	0.122883955
2015	DEC	0.794676	0	0.217285	0.578743	0.141284347
2016	JAN	0.845699	0	0.189769	0.732823	0
2016	FEB	0.934986	0.024873	0.158206	1	0.00640947

2016	MAR	0.977074	0.024873	0.156281	0.999931	0.010334887
2016	APR	0.974577	0.289182	0.306722	0.917361	0.041523553
2016	MAY	0.84788	0.243182	0.307142	0.877545	0.053269136
2016	JUN	0.835191	0.562781	0.350781	0.77114	0.096172718
2016	JUL	0.723655	0.597543	0.360957	0.828549	0.010334887
2016	AUG	0.733836	0.252322	0.350803	0.765581	0.013432287
2016	SEP	0.770946	0.041804	0.294109	0.734161	0.007390824
2016	OCT	0.781586	0.002248	0.280186	0.622752	0.013370952
2016	NOV	0.849468	0	0.197821	0.608741	0.10233685
2016	DEC	0.853213	0	0.128792	0.525989	0.141284347
2017	JAN	0.889692	0.089901	0.20245	0.692286	0.039622179
2017	FEB	0.917792	0.040306	0.169242	0.896819	0.047105005
2017	MAR	0.925485	0.129607	0.215138	0.923239	0.05173577
2017	APR	0.941891	0.264309	0.302623	0.872511	0.071454858
2017	MAY	0.842911	0.126461	0.304416	0.916024	0.080869725
2017	JUN	0.882405	0.51843	0.358578	0.86527	0.110617026
2017	JUL	0.772376	0.43602	0.379017	0.8857	0.049282385
2017	AUG	0.704022	0.227749	0.329417	0.847884	0.051797105
2017	SEP	0.798372	0.084507	0.302279	0.794673	0.10233685
2017	OCT	0.816491	0.031915	0.242419	0.6456	0.141284347
2017	NOV	0.825297	0	0.177297	0.578226	0.142204367
2017	DEC	0.834791	0	0	0	0.145240432

Appendixes 13: South Gondar Normalized Dataset all Attribute in one (2006-2017)

Year	month	T max	RR	Humidity	T min	Monthly Malaria
2006	JAN	0.626212	0.005012	0.08361987	0.310272	0.018938068
2006	FEB	0.696224	0.000323	0.05604347	0.476778	0.011296933
2006	MAR	0.705541	0.000162	0.03568218	0.571485	0.018855683
2006	APR	0.731062	0.010833	0.05373728	0.713939	0.095617161
2006	MAY	0.654827	0.244462	0.21052632	0.824528	0.194375219
2006	JUN	0.571221	0.364592	0.26315789	0.780488	0.060130167
2006	JUL	0.45706	0.911722	0.3382888	0.779591	0.036763949
2006	AUG	0.434962	0.588682	0.33657674	0.770459	0.71960538
2006	SEP	0.474462	0.341148	0.26006192	0.711704	1
2006	OCT	0.563733	0.248504	0.16087657	0.73225	0.380409038
2006	NOV	0.555779	0	0.10150376	0.531426	0.686157395
2006	DEC	0.531143	0.005982	0.09539028	0.302341	0.011276337
2007	JAN	0.536389	0	0.15231627	0.210785	0.00985521
2007	FEB	0.641485	0	0.17522272	0.29533	0.032407884
2007	MAR	0.744809	0.001778	0.11358092	0.457099	0.011822133
2007	APR	0.726791	0.047211	0.12010172	0.575875	0.087821556
2007	MAY	0.720326	0.026192	0.14204391	0.641895	0.246596503
2007	JUN	0.535245	0.461762	0.24209611	0.500138	0.06871872
2007	JUL	0.429716	0.508973	0.32309426	0.461184	0.0291949
2007	AUG	0.45706	0.531609	0.32609037	0.417208	0.204302515
2007	SEP	0.491711	0.32886	0.23772667	0.377718	0.18670319
2007	OCT	0.561508	0.186904	0.22358077	0.683148	0.080788005
2007	NOV	0.572535	0.018432	0.18376824	0.462145	0.052808271
2007	DEC	0.597278	0	0.12363927	0.131243	0.030399769
2008	JAN	0.633842	0.00291	0.1101568	0.33238	0.042654419
2008	FEB	0.673695	0	0.13571962	0.295687	0.011801536
2008	MAR	0.763886	0	0.08661597	0.342473	0.051665191
2008	APR	0.674387	0.168634	0.13135781	0.786944	0.065814677
2008	MAY	0.629073	0.141956	0.19340572	0.780792	0.142514366
2008	JUN	0.556436	0.283913	0.26094648	1	0.090210698
2008	JUL	0.455311	0.778496	0.31945614	0.657755	0.052674397
2008	AUG	0.487106	0.545837	0.32566235	0.647903	0.093670834
2008	SEP	0.516189	0.242846	0.28539837	0.598968	0.134234754

2008	OCT	0.540205	0.09135	0.22935898	0.595756	0.113412148
2008	NOV	0.549701	0.053517	0.17094969	0.31241	0.082559265
2008	DEC	0.561508	0	0.13312693	0.234094	0.019308797
2009	JAN	0.587103	0	0.02797791	0.069244	0.030791094
2009	FEB	0.689888	0	0.08992544	0.51509	0.050985521
2009	MAR	0.747988	0.012449	0.08126578	0.62267	0.071571273
2009	APR	0.768025	0.00485	0.06297183	0.639146	0.090107717
2009	MAY	0.771517	0.012935	0.08404788	0.707979	0.060047783
2009	JUN	0.672908	0.107195	0.16828837	0.765837	0.032140136
2009	JUL	0	0.516572	0.24326946	0.715429	0.052797973
2009	AUG	0.469619	1	1	0.775986	0.096265936
2009	SEP	0.554301	0.181245	0.27389468	0.567432	0.205950199
2009	OCT	0.558646	0.091835	0.26124609	0.558017	0.299765205
2009	NOV	0.610648	0.00485	0.20079611	0.306947	0.040728688
2009	DEC	0.566754	0	0.19222498	0.309431	0.001452022
2010	JAN	0.598073	0.021504	0.13290579	0.287442	0.01172945
2010	FEB	0.671759	0	0.08352815	0.40388	0.003037917
2010	MAR	0.729865	0	0.05313487	0.589748	0.031007353
2010	APR	0.79053	0.054972	0.08491818	0.906385	0.018031841
2010	MAY	0.687418	0.116572	0.18774878	0.856729	0.033581859
2010	JUN	0.621161	0.205821	0.25232198	0.738771	0.042407266
2010	JUL	0.44609	0.659337	0.32696854	0.762529	0.01178094
2010	AUG	0.441957	0.726435	0.34556505	0.755656	0.028855066
2010	SEP	0.495326	0.294584	0.31003981	0.708476	0.000978312
2010	OCT	0.580267	0.088278	0.23727725	0.661119	0.018031841
2010	NOV	0.573521	0.002425	0.18266254	0.394107	0.007188021
2010	DEC	0.542589	0	0.16002054	0.103608	0
2012	JAN	0.561508	0.021504	0.16130459	0.142778	0.000741458
2012	FEB	0.709953	0	0.0266633	0.144743	0.022109859
2012	MAR	0.70411	0	0.09453425	0.511889	0.011286635
2012	APR	0.783302	0.054972	0.06236179	0.832138	0.04150104
2012	MAY	0.67772	0.116572	0.17506749	0.779591	0.059337219
2012	JUN	0.591263	0.205821	0.26288937	0.739575	0.02153317
2012	JUL	0.484563	0.659337	0.30618767	0.674817	0.011616172
2012	AUG	0.466598	0.726435	0.33978685	0.649825	0.090519638

2012	SEP	0.499432	0.294584	0.31203008	0.624296	0.073836838
2012	OCT	0.573113	0.088278	0.21994265	0.56764	0.011533788
2012	NOV	0.548223	0.002425	0.19416188	0.444515	0.032140136
2012	DEC	0.542272	0	0.13954003	0.112499	0.021574362
2013	JAN	0.617627	0	0.0846899	0.263171	0.020977077
2013	FEB	0.726734	0	0.01690501	0.330109	0.021440488
2013	MAR	0.773266	0.045918	0.04488451	0.592661	0.011822133
2013	APR	0.753075	0.020857	0	0.565507	0.031584042
2013	MAY	0.774219	0.166532	0.10138249	0.818342	0.04905979
2013	JUN	0.648596	0.273242	0.20676692	0.809044	0.04129508
2013	JUL	0.474706	0.671625	0.30618767	0.74018	0.021327209
2013	AUG	0.463419	0.50574	0.31623176	0.697302	0.030997055
2013	SEP	0.505511	0.232821	0.29478107	0.67545	0.032140136
2013	OCT	0.578041	0.061277	0.21052632	0.617624	0.062107388
2013	NOV	0.578778	0.045432	0.2158337	0.61486	0.045002369
2013	DEC	0.605068	0	0.16644077	0.350899	0.031810599
2013	JAN	0.639566	0	0.10373657	0.256924	0.041985047
2013	FEB	0.751139	0	0.04419663	0.358118	0.032335798
2013	MAR	0.799703	0.002264	0.04124638	0.75153	0.050707474
2013	APR	0.831107	0.002264	0.02675807	0.640263	0.060140465
2013	MAY	0.770563	0.14228	0.15231627	0.877876	0.062797355
2013	JUN	1	0.240259	0.23065015	0	0.061242354
2013	JUL	0.532733	0.960388	0.31196587	0.740901	0.041377464
2013	AUG	0.481701	0.56637	0.34599307	0.732491	0.062282454
2013	SEP	0.58847	0.222959	0.29212738	0.702185	0.070428192
2013	OCT	0.589011	0.273403	0.26381418	0.713266	0.060099273
2013	NOV	0.614426	0.026839	0.18553737	0.579351	0.042396968
2013	DEC	0.592826	0	0.11186886	0.239174	0.004654707
2014	JAN	0.633048	0	0.09624631	0.312194	0.011616172
2014	FEB	0.672463	0	0.07405067	0.259353	0.020956481
2014	MAR	0.723029	0.106548	0.10801672	0.615221	0.021255123
2014	APR	0.722356	0.10768	0.1377709	0.735046	0.026465924
2014	MAY	0.689167	0.264673	0.23620722	0.74018	0.028216588
2014	JUN	0.67143	0.28844	0.24635117	0.652605	0.002821659
2014	JUL	0.598867	0.611803	0.30276355	0.649585	0.008979878

2014	AUG	0.561349	0.777365	0.31175187	0.641507	0.015828064
2014	SEP	0.55118	0.420372	0.29212738	0.728093	0.018000947
2014	OCT	0.577723	0.189814	0.24549514	0.77887	0.040368257
2014	NOV	0.609826	0	0.16147722	0.604878	0.06209709
2014	DEC	0.601538	0.000647	0.13947582	0.361763	0.030482154
2015	JAN	0.594416	0	0.12137079	0.292729	0.032212222
2015	FEB	0.768916	0.001293	0.06064005	0.454972	0.03242848
2015	MAR	0.791071	0.000647	0.04959267	0.715429	0.044466871
2015	APR	0.814186	0	0.03206546	0.766582	0.050666282
2015	MAY	0.738608	0.22118	0.19062362	0.878116	0.058595761
2015	JUN	0.666666	0.144382	0.25033171	0.842319	0.070747431
2015	JUL	0.598708	0.488601	0.28350287	0.774544	0.071571273
2015	AUG	0.586626	0.402425	0.29677134	0.762048	0.072714353
2015	SEP	0.611469	0.362005	0.28814684	0.765713	0.073414619
2015	OCT	0.651648	0.188682	0.23042901	0.779869	0.090931559
2015	NOV	0.645639	0.019725	0.20389208	0.688611	0.198216382
2015	DEC	0.610155	0.051415	0.23042901	0.581338	0.011616172
2016	JAN	0.628596	0	0.15437075	0.296518	0.031244207
2016	FEB	0.719257	0	0.09408409	0.435173	0.036763949
2016	MAR	0.833677	0.03848	0.07256638	0.650546	0.048812638
2016	APR	0.788559	0.013743	0.0584697	0.716372	0.060130167
2016	MAY	0.6871	0.276799	0.21840179	0.85745	0.060140465
2016	JUN	0.64186	0.402264	0.2576736	0.77279	0.04139806
2016	JUL	0.538297	0.662247	0.33208258	0.70846	0.04150104
2016	AUG	0.535902	0.42021	0.3126935	0.741254	0.051284164
2016	SEP	0.585513	0.443654	0.29977886	0.714932	0.061510102
2016	OCT	0.638612	0.169442	0.23800488	0.723872	0.062787057
2016	NOV	0.64761	0.000808	0.12551968	0.319115	0.082559265
2016	DEC	0.661187	0	0.09551868	0.23097	0.001410829
2017	JAN	0.679151	0	0.04535532	0.110529	0.009793422
2017	FEB	0.660318	0.041876	0.15408795	0.634282	0.020781414
2017	MAR	0.743378	0.004204	0.07625801	0.63805	0.0291949
2017	APR	0.780674	0.067745	0.11061477	0.903901	0.031285399
2017	MAY	0.660551	0.22118	0.22452241	0.903108	0.031532552
2017	JUN	0.65106	0.080841	0.2412207	0.814259	0.01731098

2017	JUL	0.547995	0.85481	0.30237834	0.793528	0.026126089
2017	AUG	0.521604	0.785287	0.31770127	0.77142	0.035486994
2017	SEP	0.586335	0.305901	0.28177797	0.749697	0.035517888
2017	OCT	0.631617	0.208407	0.25940563	0.781978	0.042716207
2017	NOV	0.605062	0.015521	0.18398939	0.569915	0.041923259
2017	DEC	0.585513	0	0.30340557	0.216312	0.009783124

Appendix 14: West Go jam Normalized Dataset all Attribute in one (2006-2017)