Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning

By

Fariha Shahrin
17121031
Labiba Zahin
17121047
Ramisa Rahman
17121006
A S M Jahir Hossain
13121007

A thesis submitted to the Department of Electrical and Electronic Engineering in partial fulfillment of the requirements for the degree of
Bachelor of Science in Electrical & Electronic Engineering

Department of Electrical and Electronic Engineering
Brac University
September2020

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I/We have acknowledged all main sources of help.

Student's Full Name & Signature:

Fariha Shahrin 17121031	**Labiba Zahin** 17121047
Ramisa Rahman 17121006	**A S M Jahir Hossain** 13121007

Approval

The thesis/project titled "Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning" submitted by

1. Fariha Shahrin 17121031
2. Labiba Zahin 17121047
3. Ramisa Rahman 17121006
4. A S M Jahir Hossain 13121007

of Summer, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Electrical and Electronic Engineering on 5th October 2020.

Examining Committee:

Supervisor:
(Member)

_____

Dr A.K.M Abdul Malek Azad
Professor, Department of Electrical and Electronic Engineering
Brac University

Program Coordinator:
(Member)

_____

Dr A.S.M. Mohsin
Professor, Department of Electrical and Electronic Engineering
Brac University

Departmental Head:
(Chair)

_____

Dr Shahidul Islam Khan
Professor, Department of Electrical and Electronic Engineering
Brac University

Abstract

Bangladesh is predominately an agriculture-based country, which faces uncertain crop yields and inefficient farming infrastructure resulting in adverse effect in food security. Habiganj is selected as the study area because of its vulnerability to floods and drought due to its unique terrain. This paper aims to present a combinational agricultural mapping and monitoring of Habiganj with crop growth and yield prediction. Multi-spectral band images of Habiganj from Landsat 8 are processed and remote sensing indices are extracted. With options of K-means and Mask R-CNN methods, crop growth is evaluated using both Python and MATLAB. Then using two type of machine learning algorithms crop yield of Habiganj is predicted from its existing parameters and the datasets are predicted by using two type of time series model. Furthermore, comparative studies are concluded between two platforms and time series model to determine the most suited environment for this research purpose.

Keywords: Agriculture, Landsat 8, Machine learning, Habiganj, crop monitoring, crop yield prediction, K-Means, Mask R-CNN, time series model

Dedication

This paper is dedicated to our parents for their continuous support and encouragement throughout the thesis work. We would also like to dedicate this paper to our supervisor for his guidance and support for the development of the final work output.

.

Acknowledgement

We would like to express our sincere gratitude to our thesis supervisor, Dr. A.K.M Abdul Malek Azad. We would like to thank him for sharing his knowledge and guiding us. Without his guidance and support the completion of the thesis would not have been possible. Special thanks to our co supervisor, Abdulla Hil Kafi who on behalf of the BRAC Onnesha team, has guided and supported us throughout the thesis to complete our work. Special thanks to Brac University's department of Electrical and Electronic Engineering for providing us with all the required facilities. We would also like to thank Laboratory of Space Engineering and Technology, School of Engineering, Brac University for providing us with all the necessity lab facilities as well. Lastly would like to thank our parents and friends for motivating and supporting us till the end.

Table of Contents

Declaration ..... **ii**

Approval ..... **iii**

Abstract ..... **iv**

Dedication ..... **v**

Acknowledgement ..... **vi**

Table of Contents ..... **vii**

List of Tables ..... **xi**

List of Figures ..... **xii**

List of Acronyms ..... **xiv**

Chapter 1 Introduction **1**

1.1 Literature review 1

1.2 Motivation of work 2

1.3 Research objective 3

1.4 Thesis overview 3

Chapter 2 Background **6**

2.1 Precision agriculture 6

2.2 Satellite images 7

2.3 Landsat 8 9

2.4 Area of interest 10

2.5 Software used 12

2.5.1 ENVI 12

2.5.2 ArcGIS 12

2.5.3 MATLAB 12

2.5.4 Python 12

2.6 Image Segmentation 13

2.6.1 K-Mean Clustering 13

2.6.2 Mask R-CNN 13

2.7 Machine learning 14

2.7.1 Random forest regression 14

2.7.2 Linear regression 14

2.8 Time series model 14

2.8.1 LSTM 14

2.8.2 ARIMA 15

2.9 Error and accuracy evaluation 15

2.9.1 Mean absolute error (MAE) 15

2.9.2 Mean squared error (MSE) 16

2.9.3 Root Mean squared error (RMSE) 16

Chapter 3 Image Processing 17

3.1 Image acquisition 17

3.2 Shapefile 17

3.3 Image merging 17

3.4 Data extraction 17

3.4.1 Normalized Difference Vegetation index (NDVI) 18

3.4.2 Normalized Difference Salinity index (NDSI) 19

3.4.3 Normalized Difference Moisture index (NDMI) 20

3.4.4 Chlorophyll index- green (CLG) 21

3.5 Area calculation 23

Chapter 4 Crop Mapping and Monitoring 24

4.1 Image Segmentation Methods 24

4.1.1 K-Mean Algorithm 25

4.1.2 Mask R-CNN 26

4.1.3 RESULT 26

4.1.3.1 Analysis and Result of the clusters techniques 26

4.2 Healthy Vegetation 28

4.2.1 Healthy Vegetation Mapping 28

4.2.2 Healthy Vegetation Area Calculation 31

4.2.3 Maximum Vegetation Area Forecasting 33

4.3 Salinity Mapping 37

4.3.1 Result 37

Chapter 5 Crop Yield prediction 40

5.1 Crop yield datasets prediction 42

5.1.1 Crop yield datasets prediction using ARIMA 42

5.1.2 Crop yield datasets prediction using LSTM 47

Chapter 6 Comparative study 52

6.1 Comparative Study of K-mean Segmentation using MATLAB and Python 52

6.1.1 K-Mean clustering 52

6.1.2 Computational platforms 52

6.1.3 Results 54

6.1.3.1 Qualitative Result 54

6.1.3.2 Quantitative Result 56

6.2 Comparative study of crop yield parameters prediction using ARIMA and LSTM 58

Chapter 7 Conclusion 63

7.1 Summary 63

7.2 Future work 64

References 65

Appendix A 70

Code for Crop yield prediction in python 70

Code for dataset prediction using ARIMA model 72

Code for k mean clustering in MATLAB 73

Code for area calculation from k mean pixel count 73

Code for dataset prediction of parametrs using LSTM 74

Code for K-Mean Algorithm in Python for Healthy Vegetation 78

Code for area calculation for images in MATLAB 80

List of Tables

Table 1:Bands of Landsat 8. Source [19] 10

Table 2: Area calculation of NDVI in ArcGIS 23

Table 3: Maximum vegetation area calculated with K-Mean clustering 34

Table 4: Accuracy comparison 41

Table 5: Prediction dataset using ARIMA 47

Table 6: Prediction using LSTM 51

Table 7: RMSE Of Python and MATLAB 56

Table 8:Accuracy and performance evaluation of ARIMA and LSTM 62

List of Figures

Figure 1:NDVI for February 2015 19

Figure 2: NDSI for February 2015 20

Figure 3: NDMI for February 2015 21

Figure 4: Chlorophyll index green for February 2015 22

Figure 5: Block Diagram of K-Mean Algorithm. Source: [42] 25

Figure 6: Mask R-CNN and K-mean of Habiganj 27

Figure 7:Healthy Vegetation Mapping of Habiganj from 2015-2019 30

Figure 8: Area Graph of Habiganj in February 32

Figure 9:Area Graph of Habiganj in June 32

Figure 10:Forecast of maximum vegetation using LSTM 35

Figure 11:Forecast with predicted data 36

Figure 12:Forecast with observed data 36

Figure 13:Salinity Mapping of Habiganj from 2015-2019 38

Figure 15:predicted yield rate for 2022 using linear regression 41

Figure 16: predicted yield rate using random forest 42

Figure 17: NDVI prediction using ARIMA 43

Figure 18: Soil moisture prediction using ARIMA 44

Figure 19: Precipitation rate prediction using ARIMA 45

Figure 20:chlorophyll index prediction using ARIMA 45

Figure 21: Crop production area prediction using ARIMA 46

Figure 22: Crop production prediction using ARIMA 46

Figure 23: NDVI prediction using LSTM 48

Figure 24: Soil moisture prediction using LSTM 48

Figure 25: Crop production prediction using LSTM 49

Figure 26: Chlorophyll prediction using LSTM 49
Figure 27: Area prediction using LSTM 50
Figure 28: Precipitation prediction using LSTM 50
Figure 29: Python Vs MATLAB. Source [51] 53
Figure 30:Difference in vegetation in MATLAB and python 55
Figure 31: Quantitative Analysis 57
Figure 32: NDVI comparison in ARIMA, LSTM and ArcGIS 59
Figure 33: NDMI comparison in ARIMA, LSTM and ArcGIS 59
Figure 34:CLG comparison in ARIMA, LSTM and ArcGIS 60
Figure 35:Precipitation comparison in ARIMA, LSTM and NASA 60
Figure 36: Crop production Area comparison in ARIMA, LSTM and Gov. 61
Figure 37: Crop production comparison in ARIMA, LSTM and Gov. 61

List of Acronyms

NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NDSI	Normalized Difference Salinity Index
NDMI	Normalized Difference Moisture Index
CLG	Chlorophyll index Green
ML	Machine Learning
PA	Precision Agriculture
R-CNN	Region based Convolutional Neural Network
LSTM	Long Short-Term Memory
ARIMA	Auto Regressive Integrated Moving Average
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
AR	Auto Regression
MA	Moving Average
ROI	Region of Interest

Chapter 1 Introduction

The agricultural sector is an indispensable part of Bangladesh's economy. This primary sector contributes a major part in the national GDP of the country and is constituent of a high percentage of the total labor force [1]. Bangladesh has a long history of degrading arable lands, floods, saline intrusion and drought threatening the sector's sustainability [2]. The country is also frequently affected by flood, drought, cyclone as the country is prone to climate change that results in loss of soil fertility and reduction in crop. We have chosen Habiganj to analyze the agricultural condition of the land. Habiganj is an ideal place for cultivation, but due to the location's unique terrain condition, the farmers face many difficulties. Thus, the purpose of this study is to provide farmers and researchers, a comprehensive analysis of agricultural factors and crop yield allowing to take effective actions for better crop productivity [3].

1.1 Literature review

Crop mapping and monitoring is an efficient way to deal with numerous issues of agricultural sector, which hinders its overall efficiency and productivity. However, to get a detailed background on this process, the paper "Crop condition and yield prediction at the field scale with geospatial and artificial neural network applications," is chosen [4]. This paper gives a detailed information regarding agricultural land use and land cover mapping. In this paper different methods for crop yield monitoring and yield predictions are explained and analyzed. This paper provides information of satellite data, approaches used to process satellite images and the methods used for agricultural land use mapping and yield prediction. A brief idea about the combination of parameters that can be used for prediction analysis is attained. To train data machine learning algorithms are used. It also gives a detailed idea about different machine learning techniques and gives us an insight regarding how accuracy evaluations can be done for the machine learning algorithms. Thus, this paper proved to be extremely helpful for this

thesis. Secondly, the paper" Predictive ability of machine learning methods for massive crop yield prediction" [5] gives a brief idea about the machine learning algorithms that can be used for yield prediction. Crop yield prediction allows early detection of problems that reduces overall crop production and, in this paper, a detailed explanation about machine learning models and algorithms that are needed for yield predictions of crops is given. The paper also compares the accuracy of different algorithms for yield prediction. For image analysis and processing, image segmentation is an important step. Image segmentation is used to segment an image into regions and extract information from it. "An improved K-means clustering algorithm in agricultural image segmentation" [6] gives an adequate idea about the segmentation process. Image stacking is quite a complicated process with processing and analyzing geospatial imagery and involves mathematical formulation to get the desired outputs. However, a detailed background on this process is accessible through the paper titled "Detection of potential arable land with remote sensing and GIS" [7]. In this paper, methods to determine potential arable lands are developed. This paper discusses how the image stacking takes place and the math regarding the band calculations that are required to extract data from satellite images. The basic information on all the band required to find out the vegetation, water, salinity parameters etc. in a region is obtained. It also shows how accurate the results will be and the type of images are expected. As this paper is able to give an idea on this thesis's result, therefore, this paper can be used as a reference.

1.2 Motivation of work

One of the basic requirements in an agricultural system is to have proper knowledge on soil characteristics and plant phenology. Precision agriculture ensures optimum productivity by using information technology. Precision farming makes use of satellite technology for crop monitoring and better yields. Throughout literature review, we have seen how satellite images can be used to extract vegetable index through images processing. Satellite images are being

used in many parts of the world to improve the agricultural sector. Extraction of soil characteristics data from the satellite images would improve productivity and farmers would be able to overcome the agricultural difficulties and produce more crops with minimal production cost. Use of satellite images to analyze agricultural field will boost the agricultural sector of the country, thus it is considered as an interesting field to be worked on.

1.3 Research objective

a) To have a clear concept on satellite images and about the bands.
b) Use the satellite image to extract soil characteristics data such as NDVI, NDSI etc. through image processing.
c) Crop mapping and monitoring for crop growth analysis and identifying the numerous issues that hinders crop growth.
d) Crop yield prediction for early detection of problems that reduces crop growth.

1.4 Thesis overview

This thesis is divided into seven chapters consisting of basic description of satellite images, image processing, mathematical equations needed and the results obtained in the research. This paper gives a detailed overview of the research with future scopes available.

Chapter 1

This chapter gives a brief idea about reason of this research, problem statement, motivation for choosing this research, goals and a brief literature review.

Chapter 2

In this chapter, we have discussed about precision farming, satellite images, Landsat 8 OLI images, our reason for choosing Landsat 8. Also discussed about our site of research. Moreover, given a brief idea about the machine learning algorithms that has been used.

Chapter 3

In chapter 3 the image processing steps are discussed. Here we have explained about the image acquisitions, the steps needed for image stacking. Also, the extracted indices from the satellite images have been explained. Moreover, the mathematical equations required for band calculations are given.

Chapter 4

In this chapter crop monitoring and mapping is explained. For crop mapping and monitoring two important factors, healthy vegetation and salinity area highlighted that plays a key part in crop growth and production. The algorithms that have been used for the mapping and monitoring also explained.

Chapter 5

In this chapter crop yield prediction and crop yield datasets prediction are explained which are considered as significant factors that contributes in crop production management. The algorithms and models that have been used and the results that have been obtained are explained in this chapter.

Chapter 6

In this chapter comparison between alternating machine learning models and software have been done. For accuracy purpose two types of programming platforms and models are being used in this study. A comparative study has been conducted between two programming

platforms to determine the suitable one for this paper. Also, another comparative study has been concluded between two time series model for obtaining an accurate prediction result.

Chapter 7

In the last chapter, the summary of our entire work is explained. We explained how our research can contribute in the agriculture sector. Furthermore, the future work that can be done based on our paper are listed.

Chapter 2 Background

In this chapter, we have discussed about precision farming, satellite images, Landsat 8 OLI images, our reason for choosing Landsat 8 images. Also, we have given a detailed information of the Landsat 8 bands. We have discussed about the area we have selected for our research and the reason for choosing the area. We have also briefly discussed about the software we have used, image segmentation and machine learning algorithms.

2.1 Precision agriculture

Precision agriculture is a farming management system which uses modern technologies like satellite system, wireless weather sensors, computers, smartphones etc. [8] for agricultural purposes. Precision farming is also termed as satellite farming or site-specific crop management. In the 1980s, the concept of precision agriculture was first introduced in USA [9]. The goal of precision agriculture is to boost the production of crops, lesser the production cost and increase the land use productivity [10]. It allows to do take agricultural decisions to be taken in the right time and the correct manner resulting in a better management of the quantity and quality of agricultural harvest. In recent days, the use of scientific methods and modern technology inventions have made the agricultural works easier as the farmers are able to cope with their tasks faster using precision farming [11]. Using spatial images of satellite imagery and analyzing them allows farmers to identify the problematic areas and can take decisions regarding what steps should be taken in the targeted areas and what time will be best for applying the methods. Using satellite data, yield maps, field maps and precise farm plans are created using the computer-based applications. In remote sensing technology, satellite data are used for the determination of soil characteristics.

Satellite images are used in precision farming for supervising crops and fields and for monitoring purposes hence reducing the environmental influences of farming. Crop monitoring

gives farmers information regarding nutrients data, state of the soil of a specified region based on which crop yield predictions can be done. Precision farming is mainly an agriculture technique by which crop, soil characteristics data, moisture data and such related data are monitored and mapped. Then they are used for analysis and necessary steps are taken accordingly [12]. Bangladesh is heavily dependent on its agriculture sector and has a long history of battling with natural calamities and through precision farming proper measures can be taken. Like other countries, it is necessary Bangladesh also adopt to such precision agricultural techniques.

2.2 Satellite images

Satellites are object that orbit around earth or other planet and make routine observation for collecting information. There are two kind of satellite: natural satellite and artificial satellite. Example of a natural satellite is the moon revolving around the earth or the earth revolving around the sun. Artificial satellite are the ones launched by humans. Artificial satellites are machines that revolve around the earth and gather information. The satellites have bird's-eye view that enables them to observe huge portion of the earth at a time hence collecting data at a faster rate than ground instrument [13]. Weather forecast is made easy using weather satellites. Landsat series invention allowed global observations on atmosphere, ocean and land surface that provides essential information, which are beneficial to humankind.

Most satellite consists of two common parts, an antenna and power source. Antenna transmits and receives information and power system controls the satellites attitude. The power system can be solar or nuclear. Satellites are equipped with sensors and cameras that gather information about the earth's land surface, water surface etc. [14].

Satellite images illustrate the earth surface in different resolutions. In remote sensing, resolution is a major characteristic. There are four resolutions that have been explained below.

Spatial resolution: Spatial resolution can be defined as the smallest measurable area on ground that a detector can sense while scanning an object in a given instant time. This is primarily determined by the instantaneous field of view (IFOV) of the sensor. In viewing a digital image, the resolution refers to the pixel size in that the resolvable object has to be bigger than the pixel size. For the detection of a very tiny feature, the sensor must have very high spatial resolution. Otherwise, the object may be blurred or not distinguishable. However, there are other factors, which affect the quality of image, such as incorrect focusing, motion or the speed of the object and atmospheric conditions.

Spectral resolution: Spatial resolution refers to the size and amount of wavelength within the electromagnetic spectrum. At this range of wavelength, the sensor collects data. To extract anticipated information from the remote sensing data it is necessary to select correct spectral bands as the probability might improve.

Temporal resolution: Temporal resolution relates to the frequency of revisits of a space borne sensor to a fixed object on the ground. This frequency is also dependent on the location of the satellite on the orbit and earth rotation speed. The distant object can be observed by the high-resolution remote sensing (RS) system that is placed in the orbital path. Hence temporal resolution quality will be much better for the object on the earth surface than that of the object located in nearer orbits. Thus, the scientist and researchers prefer using RS product with higher temporary resolution.

Radiometric resolution: The radiometric resolution of a remote sensing device indicates its ability to detect different grey-scale values, which is measured in bit. Image with more number bits has larger grey-scale values. Thus, it enables spotting differences in the reflection on the ground surface.

2.3 Landsat 8

Since 1972, Landsat satellite is operating after the launch of the first Landsat satellite, Earth Resources Technology Satellite 1[15]. For our paper, we have used Landsat 8 satellite images. Landsat 8 consist of the Operational Land Imager (OLI) and the Thermal Infrared Sensors (TIRS) providing periodic coverage of the land worldwide [16]. In our thesis, images from Landsat 8 OLI have been used. Landsat 8 OLI images have a 12-bit radiometric precision resulting in better-quality calibration which allows better land cover state characterization [17]. OLI gathers data for visible, near infrared, and short wave infrared spectral bands and panchromatic band wavebands are spectrally narrower than Landsat 7 ETM+. Landsat 8 revolves around the earth every 99 minutes gathering data of the earth land surface. It consists of eleven bands. Each wavelength band can be represented in colour image and different bands combination are used to collect earth condition information. Nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. Band 8 (panchromatic) resolution is 15 meters. Band 10 and band 11 provides more accurate surface temperature and collected at 100 meters. Landsat images provide immense information and has many users. Landsat have proved to be extremely beneficial in agricultural sector, as crop production prediction is possible using the Landsat images. Classification algorithm that differentiates two specific crops can be done using red, infrared (NIR) and short infrared (SWIR) bands of Landsat. Landsat images are used for crop monitoring. Landsat provides thermal imagery that are used to monitor consumptive water use at field level [18] and the images are available publicly. Landsat satellite are equipped with moderate resolution, multi spectral sensors that provide significant information about the field characteristics such as about the availability of nutrients, soil compositions etc. that proves beneficial for the farmers and helps farmers in crop management.

Level 1 Landsat 8 images are open and free for public use. We have downloaded the level 1 images from USGS Earth explorer software for our paper. Table 1 shows Wavelength and resolution of different bands of Landsat 8 [19].

Bands	**Wavelength (micrometre)**	**Resolution (meters)**
Band 1-Coastal aerosols	0.43-0.45	30
Band 2- Blue	0.45-0.51	30
Band 3- Green	0.53-0.59	30
Band 4- Red	0.64-0.67	30
Band 5- Near infrared (NIR)	0.85-0.88	30
Band 6- SWIR 1	1.57-1.65	30
Band 7- SWIR 2	2.11-2.29	30
Band 8- Panchromatic	0.50-0.68	15
Band 9- Cirrus	1.36-1.38	30
Band 10- Thermal infrared (1)	10.6-11.19	100
Band 11- Thermal infrared (2)	11.50-12.51	100

Table 1:Bands of Landsat 8. Source [19]

2.4 Area of interest

For this research, Habiganj area is chosen. Habiganj is a district of Sylhet division located in the north east of Bangladesh. Habiganj is located at24.3750°N 91.4167°E. The area is about 2,636.58 km².

Habiganj is mainly an agricultural district [20]. Varieties of crops like Aman rice, Aus rice, Boro rice, wheat and others are produced in this area. Different types of fruits like mango,

dates, palm, lychees, coconut etc. are produced in this region. Apart from crop cultivation, fishery and livestock are source of income in this part. Tea, which is one of the cash crops of the country, is grown in huge amount in Habiganj. More than 20 percent of the tea garden is located in Habiganj area [20]. The area is famous for rice production. Despite flood risks, Habiganj has set record in hybrid rice production [21]. It is a very fertile land with ample river channels enabling irrigation and an ideal area for cultivation. Varieties of plants are grown in this area as its climate and geographical area is suitable and favourable for the crop production. The district is few meters above the sea. It experiences good amount of rainfall. the region is dominated by the rice-based cropping pattern. Despite of having favourable agricultural conditions, the farmers face difficulties in the agriculture sector in this region due to flooding, loss of arable land, deforestation, ineffective distribution of fertilizer, scarce water and other factors [21]. Habiganj faces frequent flooding especially in the monsoon season, washing away valuable soil nutrients. It is a primary reason for causing stress and difficulties in rice cultivation [22]. Coping up with bad and unpredictable weather becomes quiet challenging for the farmers in this region. In some parts, paddy fields are destroyed by accumulation of rainwater. Moreover, the extreme weather condition like lightening proved life threatening. Thus, this region needs its problem areas identified and solutions to improve their agricultural structure and increase their yield, which is why Habiganj is chosen for this research. Habiganj is an ideal area for cultivation even though the land faces loss of arable lands, flood etc. due to its geographical placement. These weather factors give the land a continuously changing agricultural environment. Therefore, this area is determined as research area for this thesis so that some possible solutions are developed for the farmers to deal with the challenging environment.

2.5 Software used

For our research, we have used ENVI, ArcGIS, MATLAB and python

2.5.1 ENVI

ENVI is a remote sensing software. It is used for analysis of image and to extract information from images by researchers.

2.5.2 ArcGIS

ArcGIS is also used for image analysis. It is a platform that manages and creates spatial data. In this research, ArcGIS 10.8 version has been utilized.

2.5.3 MATLAB

MATLAB is a renowned platform with programming features that is highly used. The MATLAB language is a matrix-based language allowing the most natural expression of computational mathematics. The language allows it to be used for data analysis, designing and implementation of algorithms to perform operations. It is commonly used by scientists and engineer because of its range of applications [23].

2.5.4 Python

Python on the other hand is free and open-source software that can be downloaded at no cost and also modify the source code as well. Python's language is very distinct and easy to learn. It also has an extensive library, which is aimed at programming in general and contains modules for networking, databases, etc. It is extensively used for its flexibility and more manageable platform. In this paper, both the platforms have used the same algorithm and equal number of clusters to ensuring an environment for comparability and results are recorded [24].

2.6 Image Segmentation

Image segmentation is the method that extracts features from an original image. It splits an image into different set that collectively cover the entire image, or specific outlines extracted from the image and is a based on similar pixels which share the same characteristics like texture or color [25]. In this paper, two methods of image segmentation are proposed.

2.6.1 K-Mean Clustering

K-Mean algorithm is an unsupervised machine learning that segments the interested regions. It classifies the given data into multiple clusters and calculates centroid as well as the position of each cluster which has the nearest centroid from the data point. After grouping, it recalculates the new centroid of each cluster and based on that, a distance is calculated between each center and data point, which then appoints them in the cluster, which have minimum distance [26].

K-Mean clustering is a quite popular clustering method, mostly used in image segmentation. Similar criterion is described between pixels where they are gathered to form clusters. Pixels grouped into clusters and by means of clustering, which tries to use the affiliation among patterns of the set by object, the patterns in clusters or groups so that pattern inside a cluster are extra analogous to each other as compare to patterns of diverse cluster. Its dependency on the initial value of K mean is, however a limitation for the model, as it is responsible for segmentation of the colors in the image [27].

2.6.2 Mask R-CNN

Mask R-CNN is a more advanced clustering technique that incorporates deep learning techniques such a Convolution neural networks. It has two stages in its implementations. First stage is responsible for anchors which are regions analyzed by region proposal network (RPN). These anchors account for the classes and bounding box for each region. Second stage creates a binary mask for each of the following region of interest thus, is extremely useful in object

identification and instance segmentation. Therefore, first step taken is reading the image and generating region of interest and second is classifying the interest areas and generate bounding boxes and masks [28].

2.7 Machine learning

Machine learning (ML) refers to a system's ability to obtain and incorporate knowledge through large-scale observations, and continuously improving itself by learning new knowledge rather than by being programmed with fixed instructions [29]. It uses computational algorithms and converts empirical data into models used for analysis and predictions.

2.7.1 Random forest regression

Random forest is an ensemble technique used for both classification and regression. Its main idea is determining the final output by combining multiple decision trees.

2.7.2 Linear regression

Linear Regression is a supervised machine learning algorithm which helps to find a linear relationship between input and output and handles linear data.

2.8 Time series model

2.8.1 LSTM

Long short-term memory network (LSTM) is a recurrent neural network (RNN), used in sequence prediction problems. This deep learning algorithm is popular for classification, future predictions based on time series dataset architecture enable us to operate on long structures during training the network and make prediction in accordance with previous data [30]. LSTM has been widely used in language modeling, video analysis, pattern recognition, stock and medical prediction. In this study, LSTM was used to forecast future values of agricultural

parameters. LSTM time series usually requires a larger number of input data than Machine learning algorithms to work efficiently. LSTM is less sensitive to the time gap than simple RNNs and that makes it suitable for our sequential time series analysis.

2.8.2 ARIMA

The Autoregressive Integrated Moving Average (ARIMA) model is a form of regression analysis that uses time-series data and statistical analysis to interpret the data and make future predictions. The ARIMA model explain data using time series data on its past values and predict future by examining the differences between values in the series instead of through actual values. It has its own lags and the lagged forecast errors. If the series is non-stationary, ARIMA model fits the time series with a differencing process, which turns non-stationary dataset into a stationary dataset. This model consists three components- Auto Regressive (AR), Integrated (I) is the, Moving Average (MA) which are expressed by p, d, q. p indicates number of lags the model has, d is the number of times observations are differenced and q is the moving average.

2.9 Error and accuracy evaluation

For checking accuracy between the algorithm's accuracy evaluation is done using various methods.

2.9.1 Mean absolute error (MAE)

Mean absolute error (MAE) is the measurement of errors between actual and predicted value. These values are of same phenomenon. As it is focused on absolute value, only positive error is always considered.

2.9.2 Mean squared error (MSE)

Mean squared error (MSE) Measures average of the square of the difference between the actual and predicted values of the data. It is calculated by the equation

$$1/\mathrm{n}\ \sum_{i=1}^{n}(\text{actual value} - \text{predicted values})\text{^}2$$

Where n is total number of observations.

2.9.3 Root Mean squared error (RMSE)

Root Mean squared error (RMSE) is same as MSE but square root value is considered for determining the accuracy of the model.

Chapter 3 Image Processing

Satellite images are one of the most beneficial tools for analyzing and detecting changes in vegetation, salinity, water index, climate changes etc. In this study, satellite images perform as the main catalyst for datasets used for crop monitoring and yield prediction. As the necessary information for datasets cannot be obtained from the images, the information needs to be extracted through image processing.

3.1 Image acquisition

Landsat 8 images are used in this study. Images from 2014 to 2019 are retrieved from USGS [31].

3.2 Shapefile

Shapefile is a vector data format that contains geographic information like location, characteristic information used in GIS. Vector tool of ENVI software is used for creating a shapefile of Habiganj.

3.3 Image merging

Only one image obtained from satellite does not contain the whole area of Habiganj. In order to access the whole area, two images are merged using raster datasets, mosaic and clip tool of ArcGIS and the shapefile of Habiganj [32].

3.4 Data extraction

Data of remote sensing indices are extracted by processing satellite images. For this process map algebra tool of ArcGIS software is being used [32]. For crop monitoring and yield prediction, some remote sensing indices data are required. Landsat 8 images consist of different bands, which provide information about the indices of specific areas using multispectral bands of the satellite images [33]. The spectral compositions of different bands provide information

about soil characteristics, salinity index, and vegetation conditions of the earth surface. In this paper, calculations have been done for Normalized Difference Vegetation index (NDVI), Normalized Difference Salinity index (NDSI), Normalized Difference Moisture index (NDMI) and Chlorophyll green (CLG) using different band calculation of the Landsat 8 image.

3.4.1 Normalized Difference Vegetation index (NDVI)

NDVI is one of the mostly commonly used vegetation indices that gives information about the distribution of plants. It shows the density of plant growth over the selected area. NDVI mainly provides information about the green plants as it is highly sensitive to green vegetation. By measuring NDVI plants, photosynthetic characteristics can be determined. NDVI is also an indicator of plant heath and density of plant growth over the selected area. For high NDVI value, the red and NIR will have low reflectance and high reflectance respectively and vice versa. The generated result lies between -1 and +1. NDVI high values indicates high vegetation areas like forests. Low values indicate barren areas [34][35].

NDVI calculation formula is-

$$(NIR - Red) / (NIR + Red)$$

It is classified into three type- high, moderate and low. Figure 1 shows the NDVI of Habiganj area for February 2015.

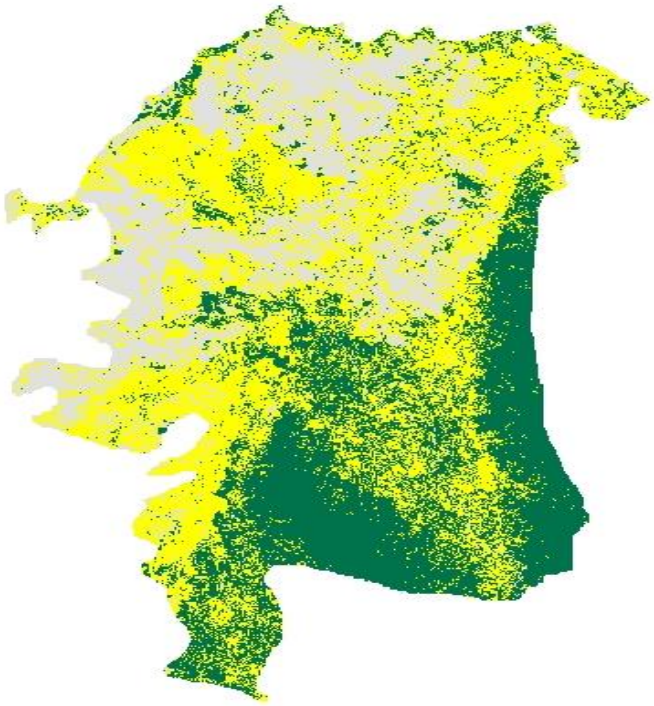


Figure 1:NDVI for February 2015

3.4.2 Normalized Difference Salinity index (NDSI)

NDSI provides information about the salinity concentration in the soil. Saline intrusion has a negative impact on the soil as it causes destruction of plants, reduces growth of the plant and other organisms in soil and reduces productivity of land [36]. An increase in soil salinity reduces the water uptake capability of plants hence resulting in water stress in plants and in nutrient imbalance and that limits the plant growth. High salinity index reduces soil productivity. In Figure 2, the NDSI of Habiganj is shown for February 2015.

$$NDSI = (red - NIR)/ (red+ NIR)$$

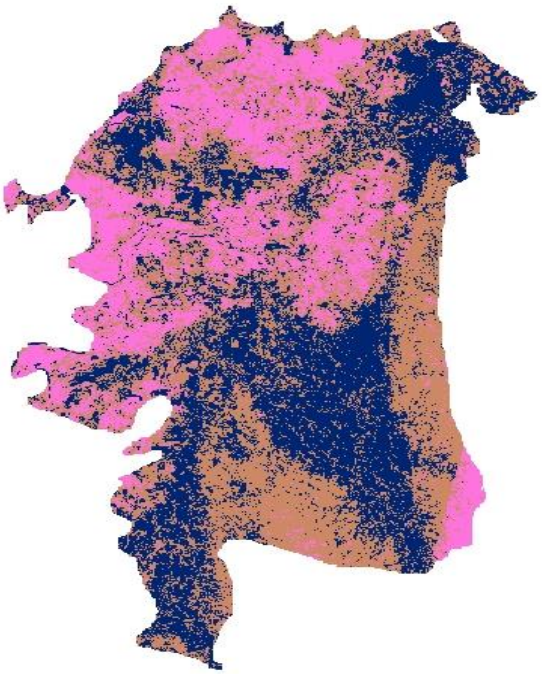


Figure 2: NDSI for February 2015

3.4.3 Normalized Difference Moisture index (NDMI)

NDMI is used to determine the water content of the vegetation. NDMI is most commonly used to determine areas to determine the moisture content present in soil and also to monitor the changes over time. NDMI value ranges from -1 to 1[37]. Extremely low moisture content reduces plant growth yield whereas excess content results in root disease of plants. Hence, a balanced amount of NDMI is required for healthy vegetation. The accuracy in retrieving the vegetation water content is improved as variations initiated by leaf's internal structure and dry matter content are removed by NIR and SWIR1. In the SWIR interval of the electromagnetic spectrum, the spectral reflectance is largely controlled by the internal leaf structure's amount of water. Hence, it is determined that SWIR reflectance has a negative relation to leaf water content. NDMI is one of the important catalysts that plays an impertinent role in crop yield

prediction. In Figure 3, the NDMI of Habiganj is shown for February 2015. Formula for NDMI is-

$$NDMI = (NIR-SWIR1)/ (NIR+SWIR1)$$

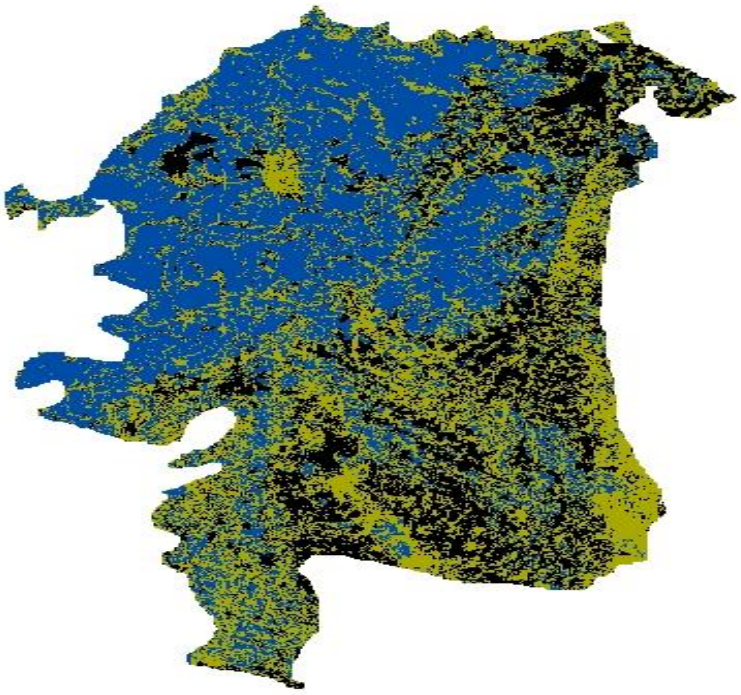


Figure 3: NDMI for February 2015

3.4.4 Chlorophyll index- green (CLG)

Chlorophyll Index-Green estimates chlorophyll content of leaf across a wide range of plant species. It is an indicator of photosynthesis activity. NIR and green wavelengths that are used in here provides a better prediction of chlorophyll content [38]. Chlorophyll is a vital component for photosynthesis of plants. In plants chlorophyll green traps sunlight that is required for the conversion of water and carbon dioxide in photosynthesis process. Nitrogen

content and chlorophyll green content are related. Nitrogen content has impact on leaf growth, affecting the arability of selected area. Plant nitrogen content and chlorophyll green index are proportional to each other [39]. Plant's chlorophyll green index can be calculated by the equation given below

$$CLG = [(NIR / Green)-1]$$

In Figure 4 the chlorophyll green index of Habiganj is shown for February 2015

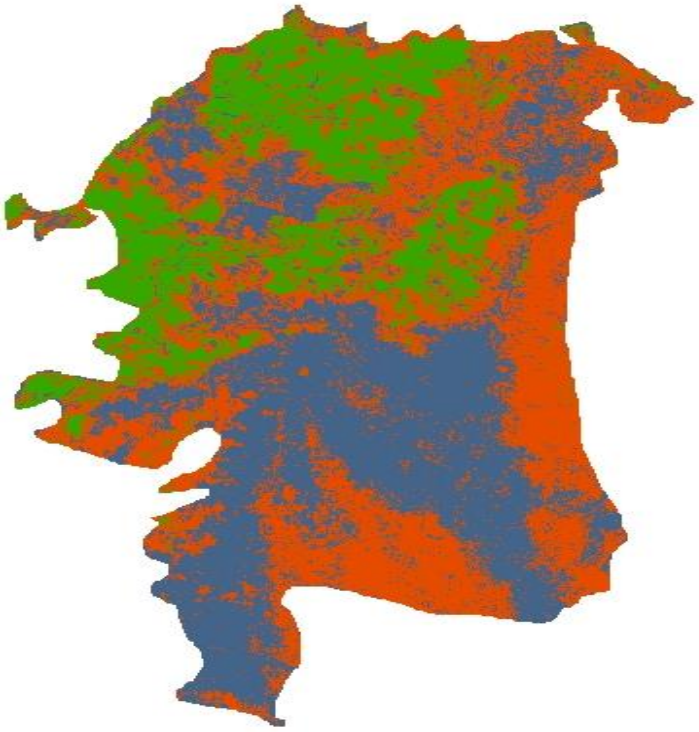


Figure 4: Chlorophyll index green for February 2015

3.5 Area calculation

Using maximum likelihood classifier tool in ArcGIS, pixel counts for high, moderate and low area are attained. Later, area calculation is done using area field calculator with obtained pixel count and spatial resolution data. Unit for area is square kilometre.

In table 2 area calculation for NDVI in February from year 2014-2019 using ArcGIS is shown

YEAR	HIGH	MODERATE	LOW
2014	699.678	1124.1855	726.7635
2015	592.947	1140.9363	816.7437
2016	605.2833	1055.1222	890.2215
2017	574.7697	1170.6876	805.1697
2018	558.27	1289.5389	702.8181
2019	708.8508	1241.0541	602.5185

Table 2: Area calculation of NDVI in ArcGIS

Chapter 4 Crop Mapping and Monitoring

Agricultural sector deals with numerous issues, which hinders overall efficiency and productivity throughout the years. Some of the problems involves inadequate land management, depleting soil resources causing limitations in the farm productivity. Developments in this sector has a considerate amount of progress and implementing remote sensing techniques is one of them.

Remote sensing equipped with its high spatial range and resolution allows extensive mapping of large stretches of land and providing information. The information is quite comprehensive like aerosol monitoring data, soil and vegetation properties as well status of cultivated crops and even future forecasts on the productivity. These combined creates a more detailed decision framework for farmers introducing a new dimension for efficiency known as precision farming.

In this paper, two important factors, healthy vegetation and salinity are highlighted which plays a significant role in crop growth and production.

Remote sensing indices such as Normalized Difference Vegetation Index (NDVI) is implemented to detect healthy vegetation and Normalized Difference Salinity Index (NDSI) is used to map the salinity areas. The areas are evaluated using two methods of segmentation presenting different accuracy levels.

4.1 Image Segmentation Methods

Habiganj Image is extracted using ArcGIS maximum likelihood, which uses a fixed threshold range for low, moderate and high vegetation. To classify the three vegetation further, image segmentation is used to separate each type of vegetation for proper vegetation mapping. In this section, two types of segmentation are used and evaluated. Firstly, K-mean, which is a traditional clustering technique and Mask R-CNN, which is a more advanced segmentation method.

4.1.1 K-Mean Algorithm

K-Mean is a popular clustering method because of its simplicity. It assembles data points according to the number of clusters defined by user. The K-mean cluster thus separates the areas of different clusters. In this paper, the clusters are adjusted according to the various pixels color which define the vegetation categories and only the high vegetation region displayed. The healthy region of Habiganj have a NDVI value over the rest of the region are then masked, converting them into black pixels thus highlighting only the healthy vegetation areas in Habiganj.

The following process has been carried out in Jupiter notebook in python language. OpenCV is used as a library applied in this paper for utilizing K-Mean algorithm. OpenCV is commonly used for computer vision purposes because of its feasibility and option for modification of its codes [41]. Using K-Mean algorithm in OpenCV, Habiganj images are uploaded and with the appropriate number of clusters, the vegetation is separated and only the healthy vegetation area is shown. Flow chart of k mean is shown in figure 5 which provides a detail on the process.

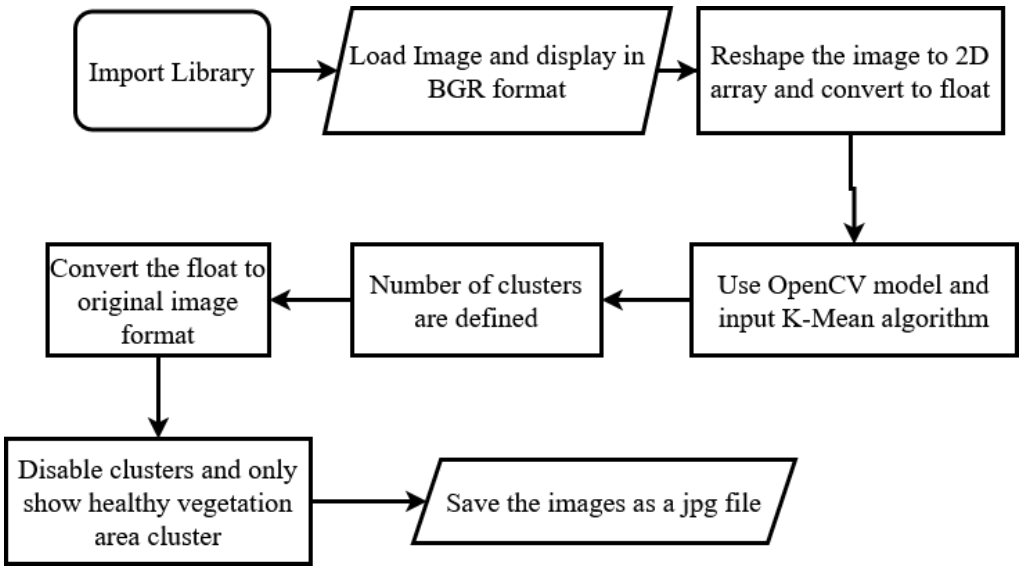


Figure 5: Block Diagram of K-Mean Algorithm. Source: [42]

4.1.2 Mask R-CNN

Mask R-CNN is a more advanced clustering technique that incorporates deep learning techniques such a Convolution neural networks. Habiganj images are divided to training and testing data while ROIs are manually generated using an online annotator tool. In this paper, retrained dataset from Microsoft COCO is used even if the healthy vegetation region did not have its own class. To perform the training, 50 epochs with 100 steps per epoch are taken.

Even with Mask R-CNN being very complex, with framework of resent 50, thus taking a considerable amount of time. The final result consisted of Habiganj with, multiple masks that indicate only healthy vegetation areas

4.1.3 RESULT

4.1.3.1 Analysis and Result of the clusters techniques

Both methods bring out instance segmentation results however with vastly different methods. With observation as shown in Figure 6, K-Mean is much more accurate with detecting the vegetation areas with more defined areas. The only limitation is manually determining the number of clusters, which affects the final output

While Mask R-CNN brings in a more automated system where after training, it takes in any Habiganj image and segments the healthy vegetation areas. However, since it is highly dependent on the number of datasets for accuracy, it does not have a uniform distribution of the area and misses out minuscule vegetation areas of Habiganj in the high-resolution satellite images. It also fails to represent the healthy vegetation area with a defined mask shape and have an overlap of different vegetation areas within the mask. Thus, for further calculation of area, K mean algorithm is chosen for crop monitoring in this paper for better accuracy. In figure 6, K-mean and Mask R-CNN of Habiganj area is shown.

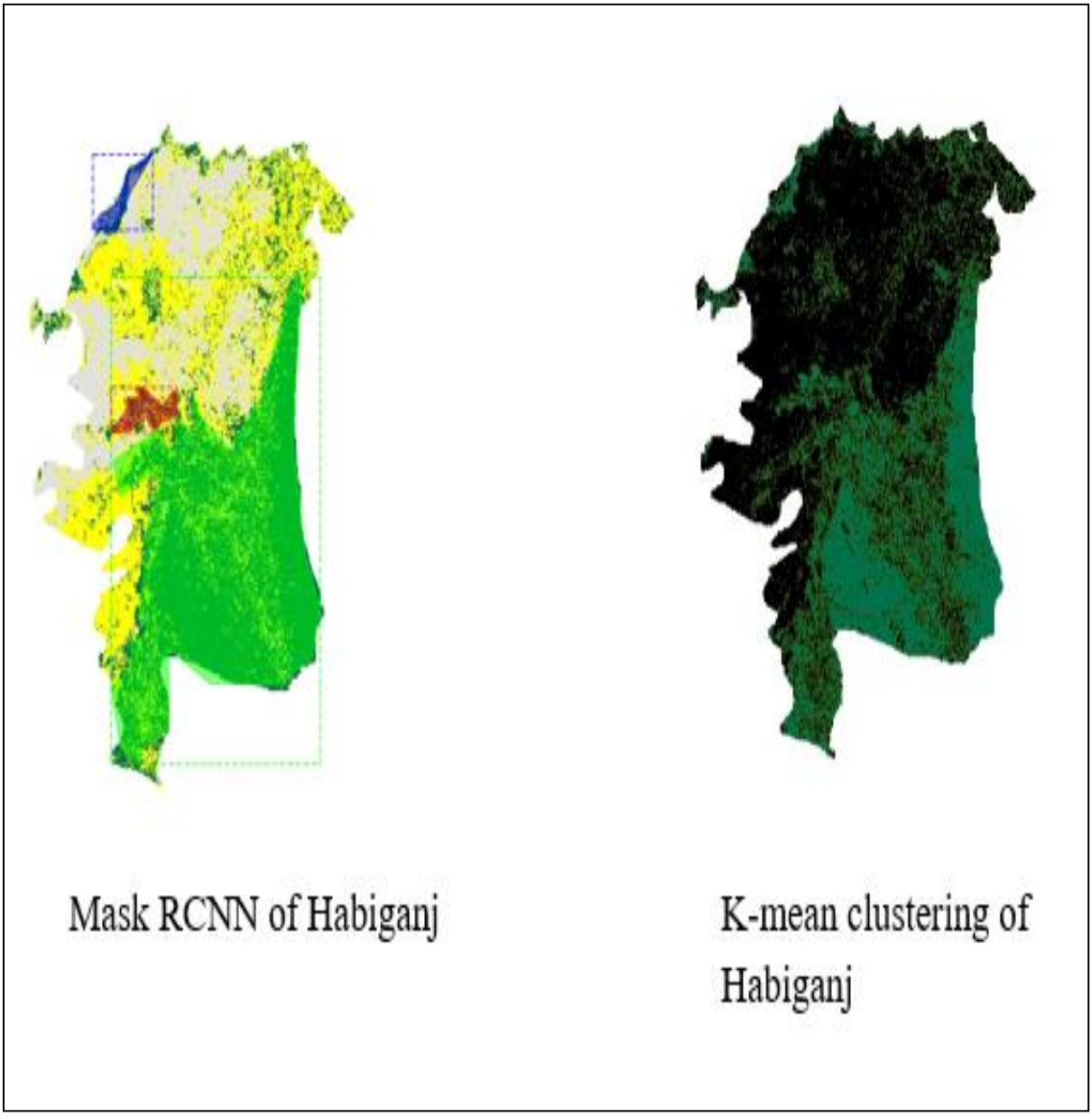


Figure 6: Mask R-CNN and K-mean of Habiganj

4.2 Healthy Vegetation

In the following section, NDVI is used to extract healthy regions from Habiganj Dataset. Healthy vegetation absorbs most of visible light emitted by the sun and reflects a large portion of the near infrared light, which is detected for NDVI calculation. Therefore, unhealthy or sparse vegetation reflects more visible light and less near infrared light. The difference in reflectance by healthy or well-stocked vegetation and that of unhealthy or sparse vegetation is calculated as

NDVI=(NIR-RED/(NIR+RED)

where NIR is the near-infrared reflectance and RED is the reflectance of visible light. NDVI is associated with leaf area index (LAI) thus indicating that NDVI is linked with chlorophyll, an essential pigment in plants that is accountable for the plant's food production [40].

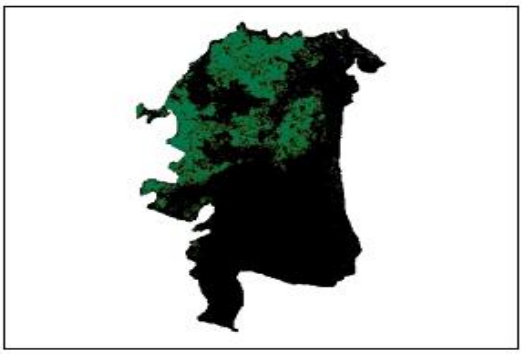
In this case, NDVI calculation is applied to Habiganj image, the preexisting vegetation is classified into three types. The types are maximum vegetation, moderate vegetation and minimum vegetation. The minimum vegetation NDVI value lies near the negative value so it is often considered barren lands with little or no vegetation at all. Moderate Vegetation consists of unhealthy vegetation as well as low-lying shrubs and grass while maximum vegetation depicts healthy crops and trees in the region.

4.2.1 Healthy Vegetation Mapping

As it has been evidently demonstrated that K-mean performs better than Mask RCNN, Habiganj images then were thus processed with the K-Mean Algorithm.Landsat-8 images obtained during February and June. February is a peak harvesting month for wheat while June is a harvesting month for rice in Bangladesh thus making these images ideal for analysis [43].

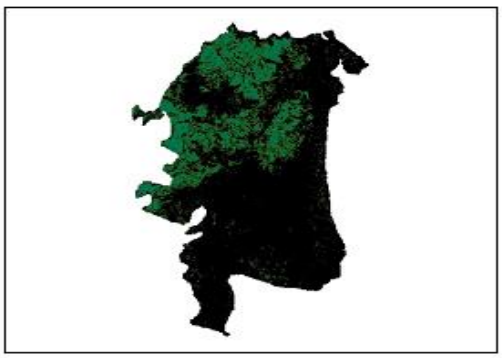
February 2015

June 2015

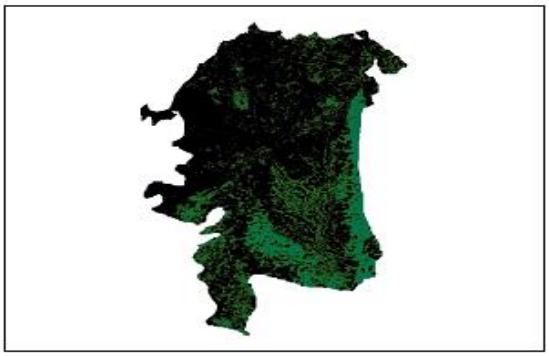


February 2016

June 2016



February 2017



June 2017

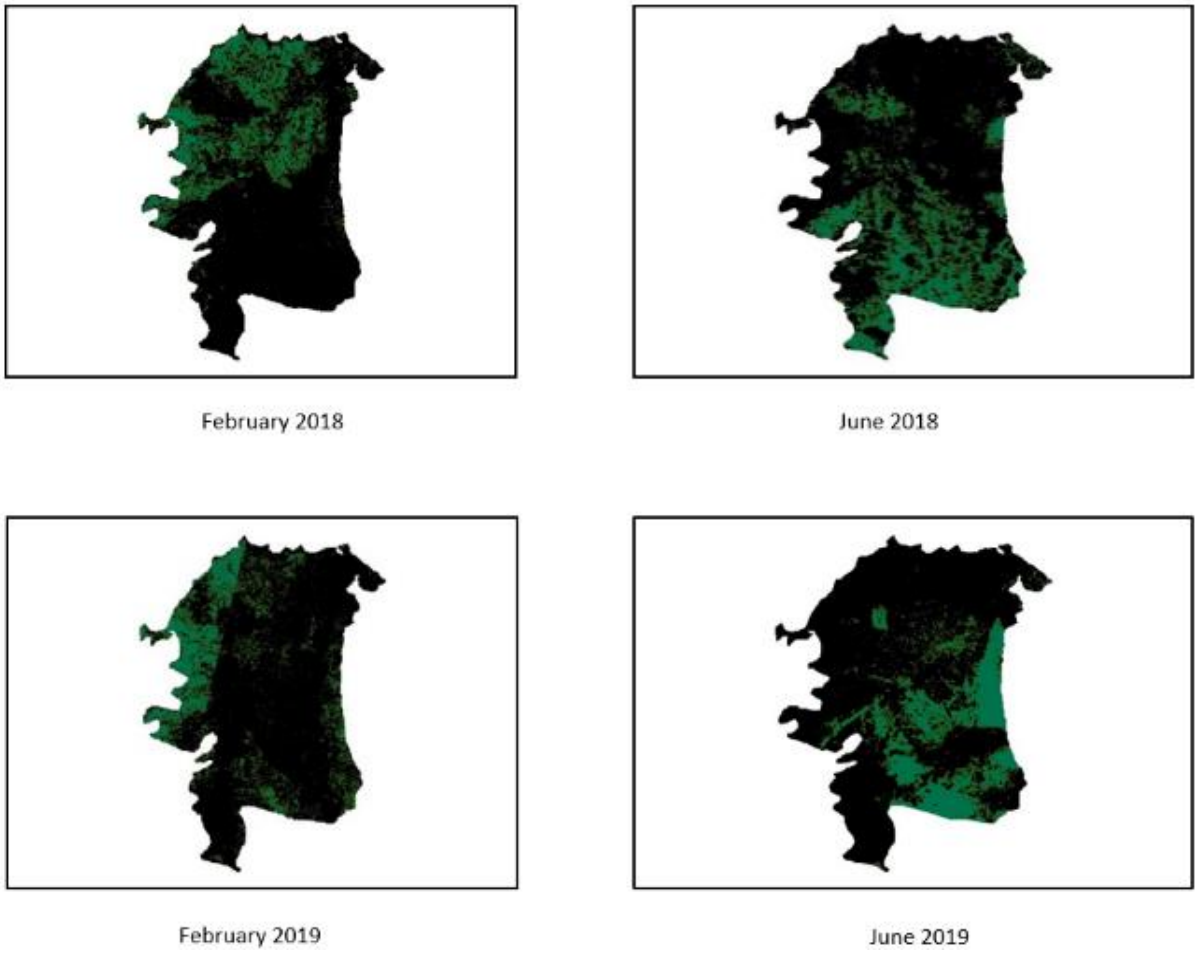


Figure 7:Healthy Vegetation Mapping of Habiganj from 2015-2019

The datasets used in this study is used from 2015 till 2019 and the data is processed using K-Mean algorithm. The Habiganj images shown in Figure 7, shows only areas of healthy vegetation as a result thus making the changes over the years quite discernible. It makes it easier to note down trends and analyze further.

However, it is noticed that during rainy seasons, especially during June-July months, due to increase in cloud coverage [44], resulting into inaccuracies in NDVI mapping thus being a limitation for Landsat-8.

4.2.2 Healthy Vegetation Area Calculation

To further analyze the healthy vegetation areas, the accumulated processed Habiganj images, is used. The area of healthy vegetation is calculated by converting the images to binary images and using relevant libraries to count the number of white pixels. Thus, using the total number of pixels of Habiganj with the healthy vegetation area pixels, an estimation of the area is calculated. The areas over the year are then visualized in a graph and the trend can be analyzed for crop growth.

In Figure 8, during February over the years, a considerable decline is noticed in vegetation area of Habiganj. February is a harvesting month for barley and other minor crops and it can be seen that the yield and production area has been decreasing over time from Agriculture Handbook from Bangladesh Bureau of Statistic. The handbook also shows an increase in the respective crops in 2019 resulting in a slight peak in the graph.

In June, it is a major harvest month for rice, crops and fruits due to the month residing in the monsoon season. As shown in Figure 9, a noticeable descent of healthy vegetation is at 2019 which indicates to a flood incident that occurred in Habiganj which was recorded in the same year [45].

Thus, it can be evidently stated that high vegetation area corresponds to the yield and production area parameters acquired from the official agricultural handbook [43]. This analysis is useful especially in case of detecting early signs in crop growth and condition.

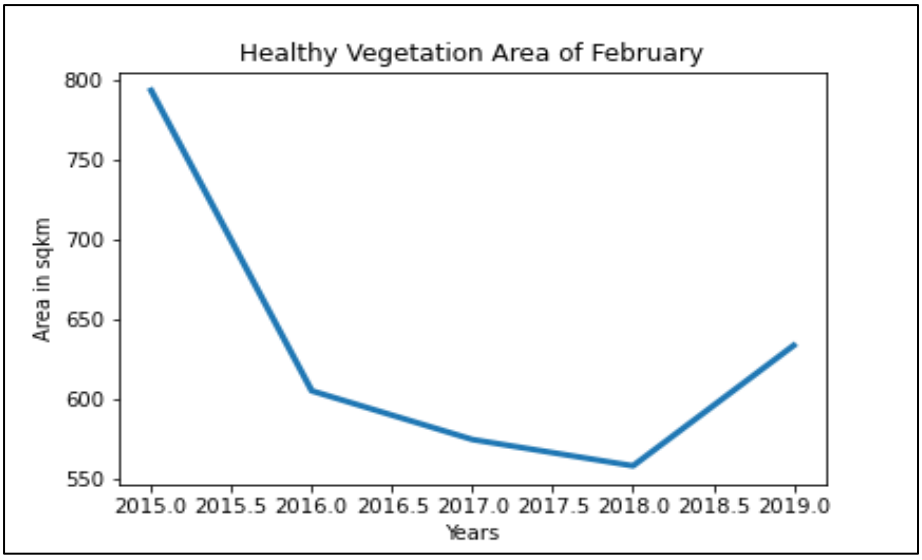


Figure 8: Area Graph of Habiganj in February

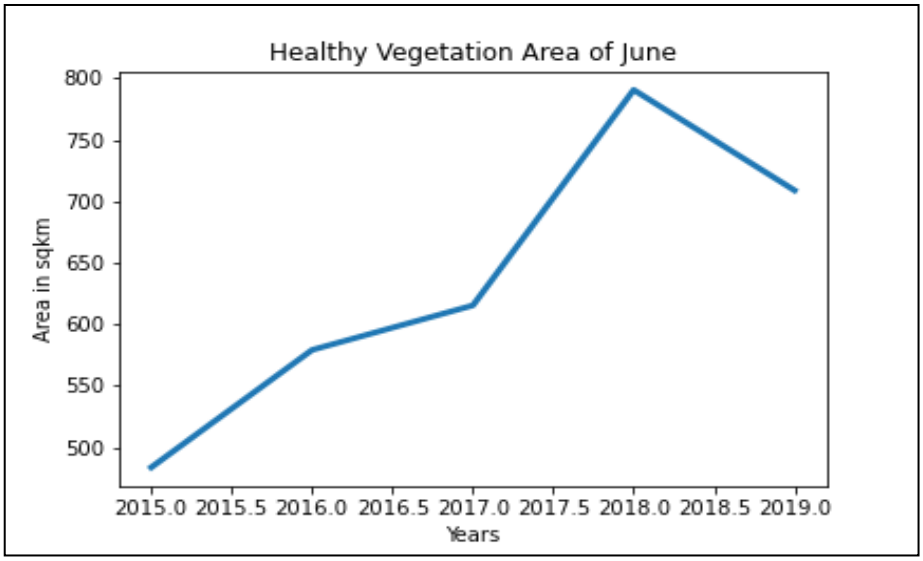


Figure 9:Area Graph of Habiganj in June

4.2.3 Maximum Vegetation Area Forecasting

February, June and November maximum area values, extracted from K-mean clustering are taken as seasonal values as spring, summer and fall respectively for the convenience of the time series model and forecasting is done. The network is trained first with 90% of the input as test data and the rest 10% is used for accuracy evaluation. It forecasts 8 seasonal steps ahead of last test data. 8 seasonal steps show almost two years of forecast. More steps can be taken but that changes the accuracy.

Later root means square error (RMSE) is calculated with both standardized data and test data to evaluate the accuracy of the Network. Table 3 shows the inputs for maximum vegetation area forecast that we extracted using K-Mean clustering.

Season	Area (Square km)
Spring 2015	707.4335
Summer 2015	519.8461
Fall 2015	931.6489
Spring 2016	653.7001
Summer 2016	609.93
Fall 2016	1370.162
Spring 2017	711.143
Summer 2017	804.285
Fall 2017	1806.013
Spring 2018	720.4814
Summer 2018	823.7945
Fall 2018	1229.641
Spring 2019	779.3832
Summer 2019	683.4217
Fall 2019	1290.473
Spring 2020	817.705
Summer 2020	476.2357

Table 3: Maximum vegetation area calculated with K-Mean clustering

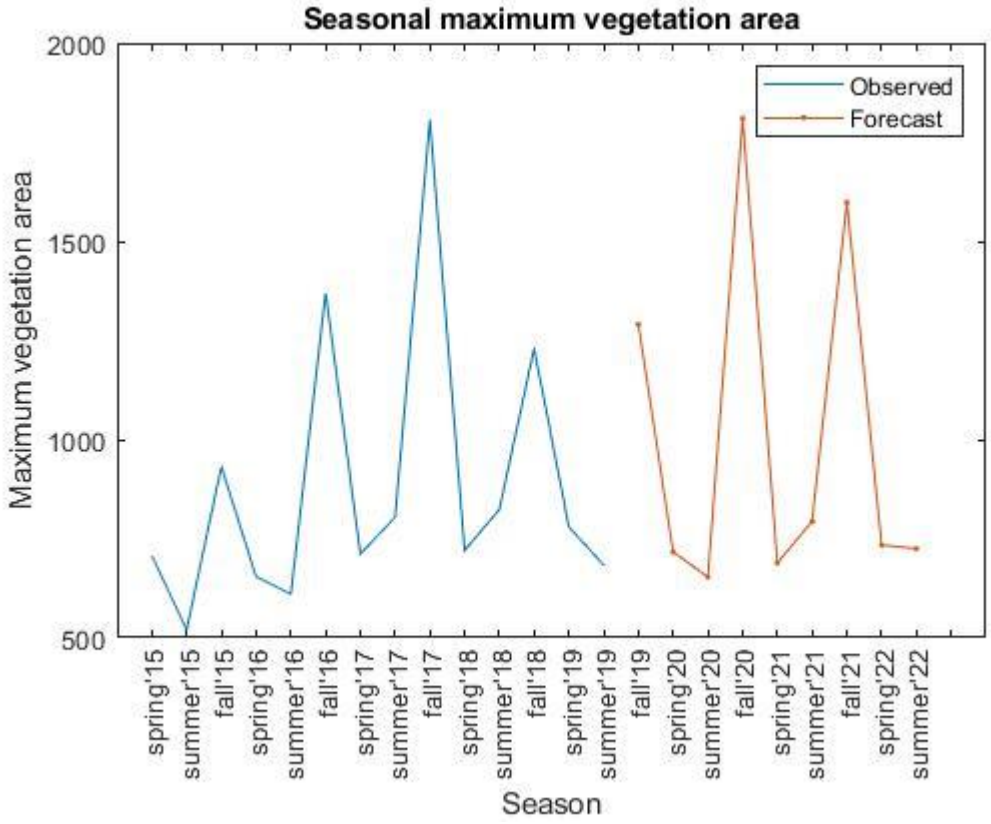


Figure 10:Forecast of maximum vegetation using LSTM

Accuracy evaluation with RMSE

Usually for a LSTM network, time series predictions are more accurate when updating the network state with the observed values instead of the predicted values. In this case the model acts the other way around because of a very small number of inputs. Forecast with predicted and observed data shown in figure 11 and 12.

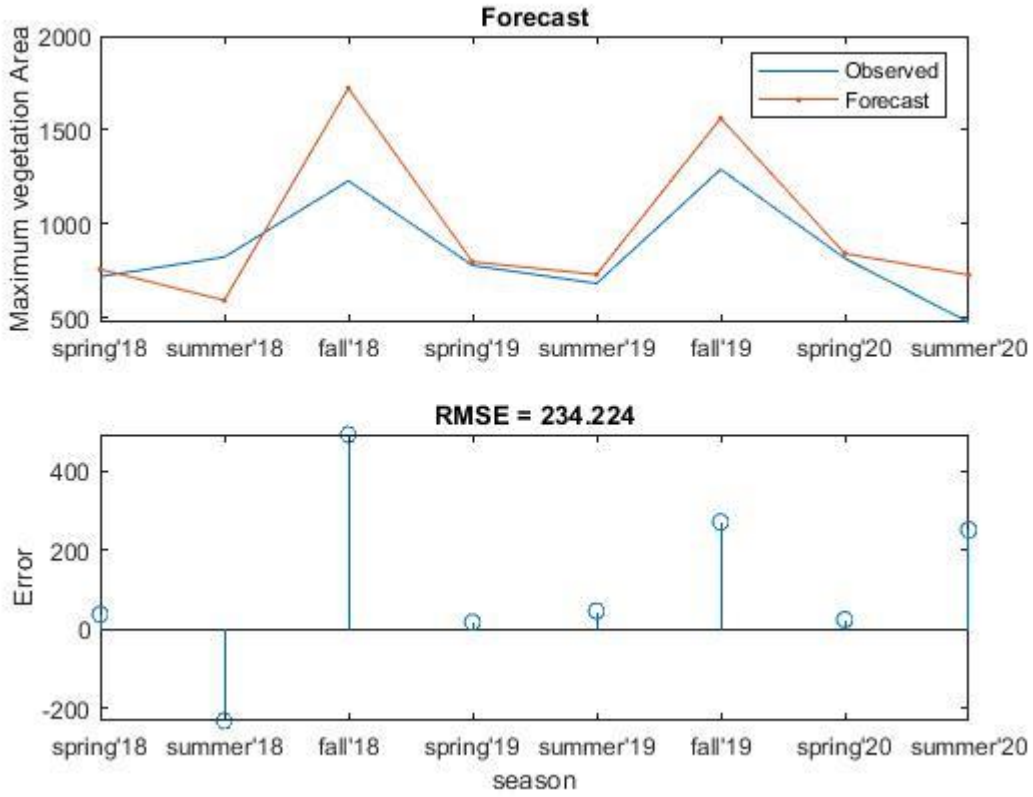


Figure 11:Forecast with predicted data

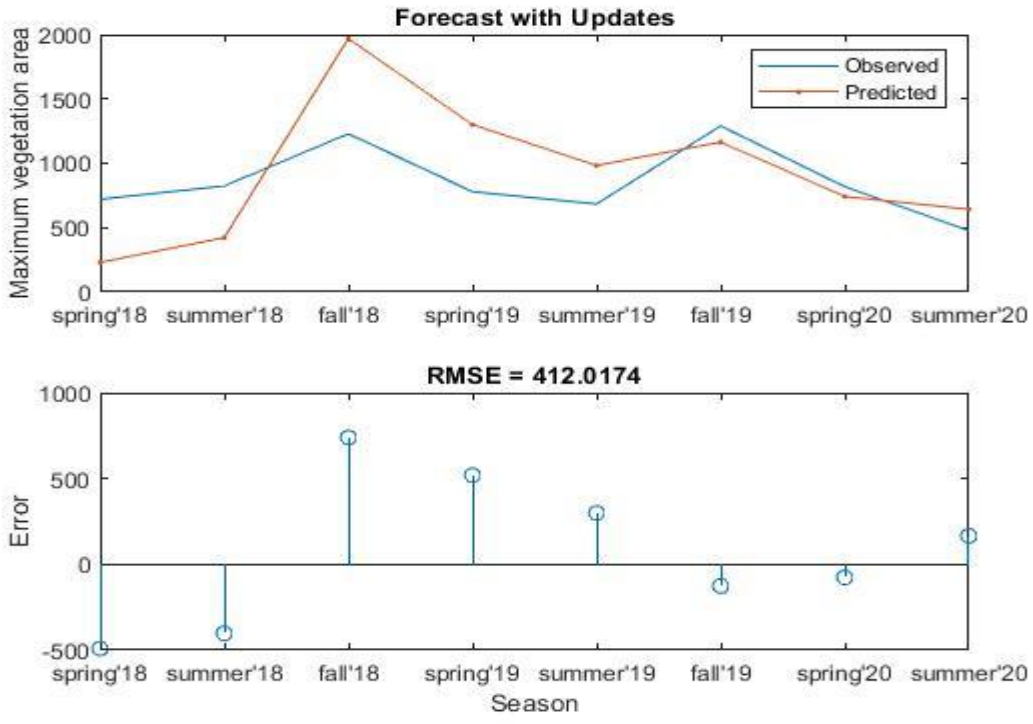


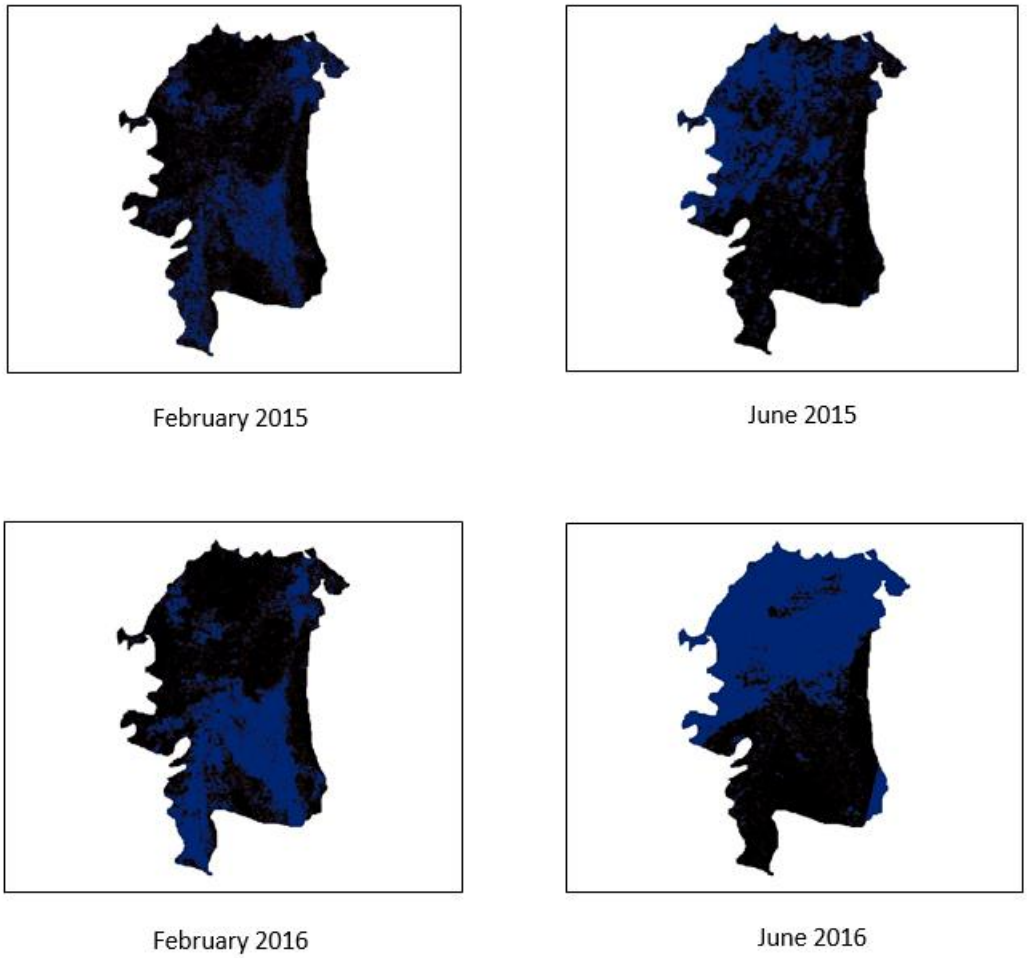
Figure 12:Forecast with observed data

4.3 Salinity Mapping

Salinity is a major factor that contributes to loss of soil productivity and yield. It is quite difficult to find out exact areas of high salinity thus increasing the risk of it encroaching to the soils with cultivated crops. Thus, in this paper, with remote sensing NDSI, high salinity areas are highlighted using K-Mean to show which areas in Habiganj needs oversight for better yield.

4.3.1 Result

Using NDSI, Habiganj image is classified to high, moderate and low salinity. The areas with high salinity are alarming as it worsens agricultural conditions and prevents adequate crop growth. By identifying the regions with high areas as shown in Figure 13, immediate salinity stress management can be implemented to reduce its adverse effects.



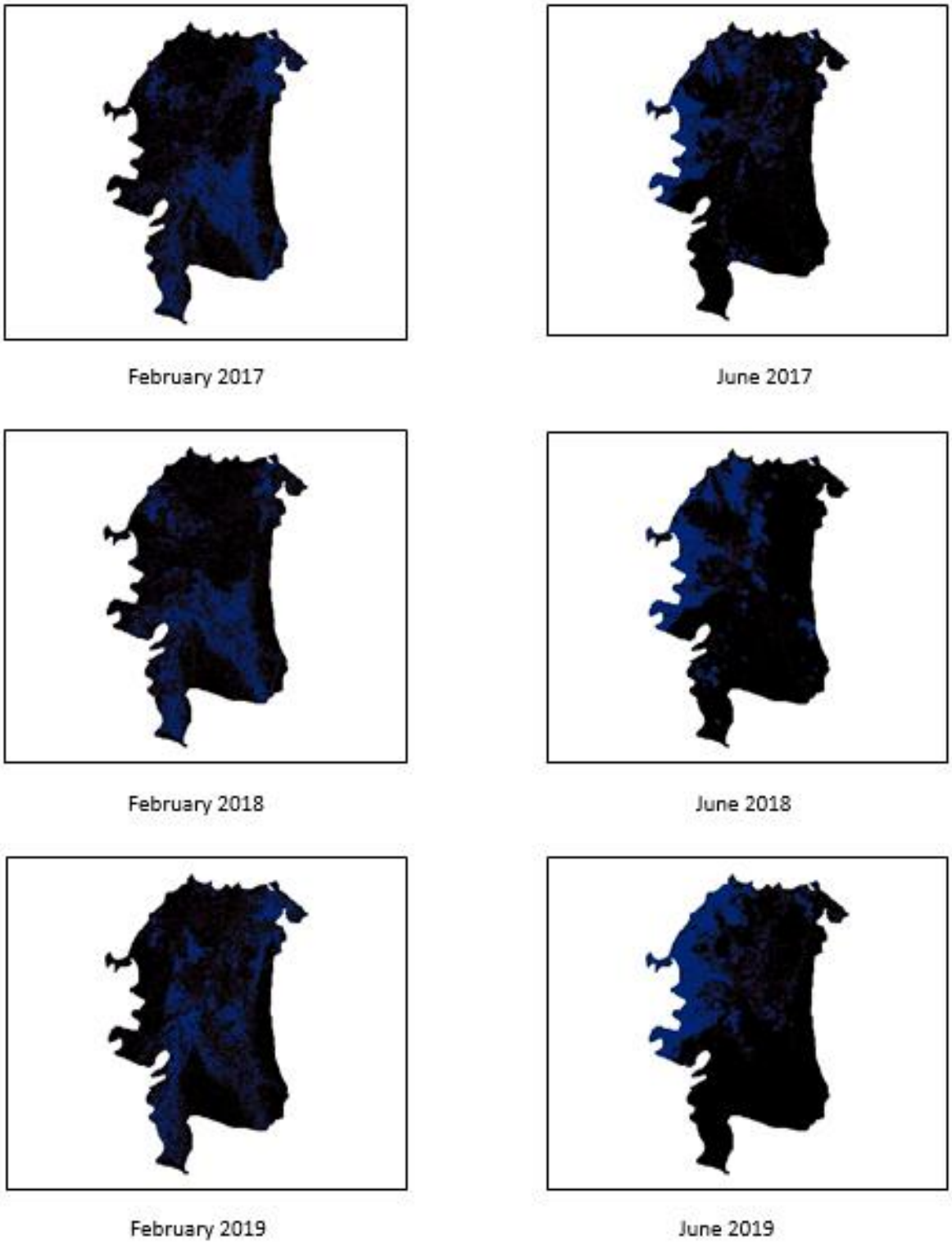


Figure 13:Salinity Mapping of Habiganj from 2015-2019

Thus, with the identification of such areas, measures can be taken to ensure proper crop growth. Some measures that can be taken for tackling salinity is introducing more alkaline to the soil or growing crops that can resist higher saline concentrations. Irrigation and fertilizer monitoring are one of the important practices for a reduction in saline concentration which would ensure future crop growth [46].

Chapter 5 Crop Yield prediction

Unpredictability of problematic situations has always been a major negative factor in reduction of crop productivity. Crop yield prediction is one of the most significant factors that contributes in crop production management. Early prediction would help in early detection of problems and help researchers and farmers to prepare for the forthcoming predicaments thus, leading to higher food security [47]. As crop production is dependent on various parameters such as NDMI (soil moisture), NDVI, precipitation rate, production area, crop production, chlorophyll Index etc. that are used as datasets in prediction. Remote sensing indices data are obtained by image processing while rest of the parameters such as crop production area, crop production and precipitation rate data are collected from Bangladesh Bureau of Statistic and NASA Giovanni website respectively. Machine learning is used for processing these datasets. Two types algorithms linear regression and random forest regression are used to predict yield for accuracy purpose [48]. Linear and random forest both regression algorithm is based on supervised data. While linear regression focuses on linearity, random forest regression focuses on both linearity and non-linearity and handles collinearity well unlike linear regression. Both algorithms have different positive side and negative side. In order to determine output linear regression focuses on finding out a linear relationship between input and output whereas random forest determines the final output using multiple decision trees. To attain more accurate result both linear regression and random forest algorithms are used for crop yield prediction. For accuracy evaluation MAE, MSE and RMSE are implemented. Table 4 shows the accuracy comparison between these two algorithms.

METHOD	Linear Regression	Random Forest Regression
MAE	92.92%	94.86%
MSE	93.49%	95.87%
RMSE	93.49%	95.87%

Table 4: Accuracy comparison

Using linear regression predicted yield rate for year 2022 is shown in Fig 15

```
#predicting the single observation results

from sklearn.externals import joblib

YEAR=2022

Soil_moisture=0.14478866
NDVI=0.24079038
Precipitation_rate=14.2975175

Area=260039.592
Production= 605634.867
Chlorophyll_index= 0.88538812
prediction_data=[YEAR,Soil_moisture,NDVI,Precipitation_rate,Area,Production, Chlorophyll_index]
prediction_data_array=np.array(prediction_data)
prediction_data_array=prediction_data_array.reshape(1,-1)
model=open("multiple_regression_model.pkl","rb")
prediction_model=joblib.load(model)
print(prediction_data_array.size)
model_prediction=pred_model.predict(prediction_data_array)
round(float(model_predictioction),2)
```

```
7
10.09
```

Figure 14:predicted yield rate for 2022 using linear regression

Using random forest regression predicted yield rate for year 2022 is shown in Figure 16

```
#predicting the single observation results

from sklearn.externals import joblib

YEAR=2022

Soil_moisture=0.14478866
NDVI=0.24079038
Precipitation_rate=14.2975175

Area=260039.592
Production= 605634.867
Chlorophyll_index= 0.88538812
prediction_data=[YEAR,Soil_moisture,NDVI,Precipitation_rate,Area,Production, Chlorophyll_index]
prediction_data_array=np.array(prediction_data)
prediction_data_array=prediction_data_array.reshape(1,-1)
model=open("multiple_regression_model.pkl","rb")
prediction_model=joblib.load(model)
print(prediction_data_array.size)
model_prediction=pred_model.predict(prediction_data_array)
round(float(model_predictioction),2)
```

```
7
10.33
```

Figure 15: predicted yield rate using random forest

5.1 Crop yield datasets prediction

Higher accuracy in prediction can be achieved by using all of the parameters as datasets. As crop yield prediction is based on various parameters' datasets, prediction of these datasets can ensure better accuracy [49] [50].

5.1.1 Crop yield datasets prediction using ARIMA

For predicting crop yield datasets, time series analysis model ARIMA and Python as the main platform are used. Input datasets is divided into training and testing datasets and model is

trained using ARIMA (1,1,0) predictions of datasets are done. Prediction of datasets are shown in figure 17,18,19,20,21 and 22.

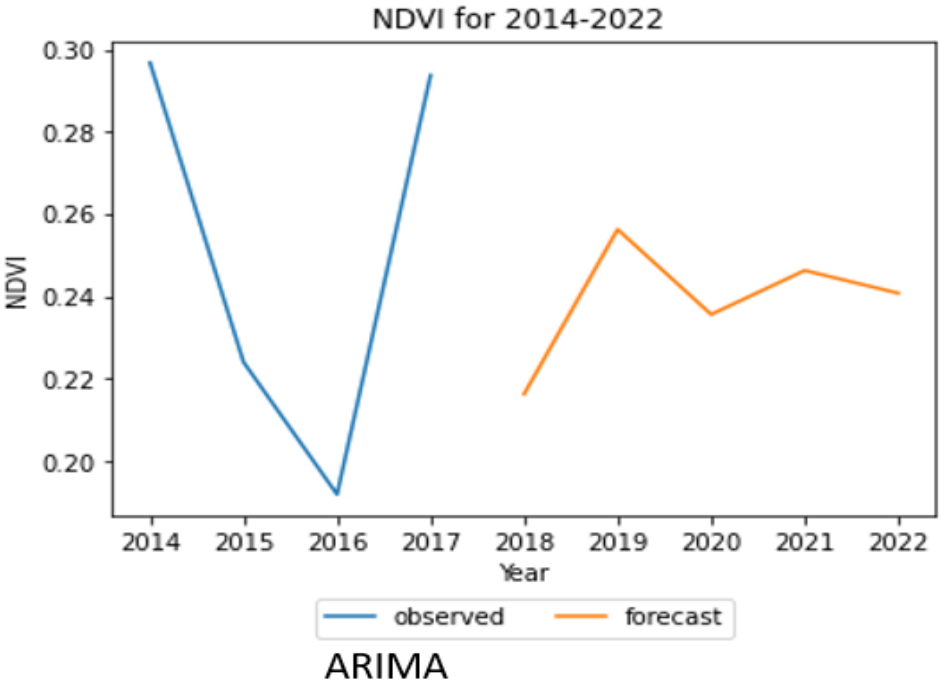
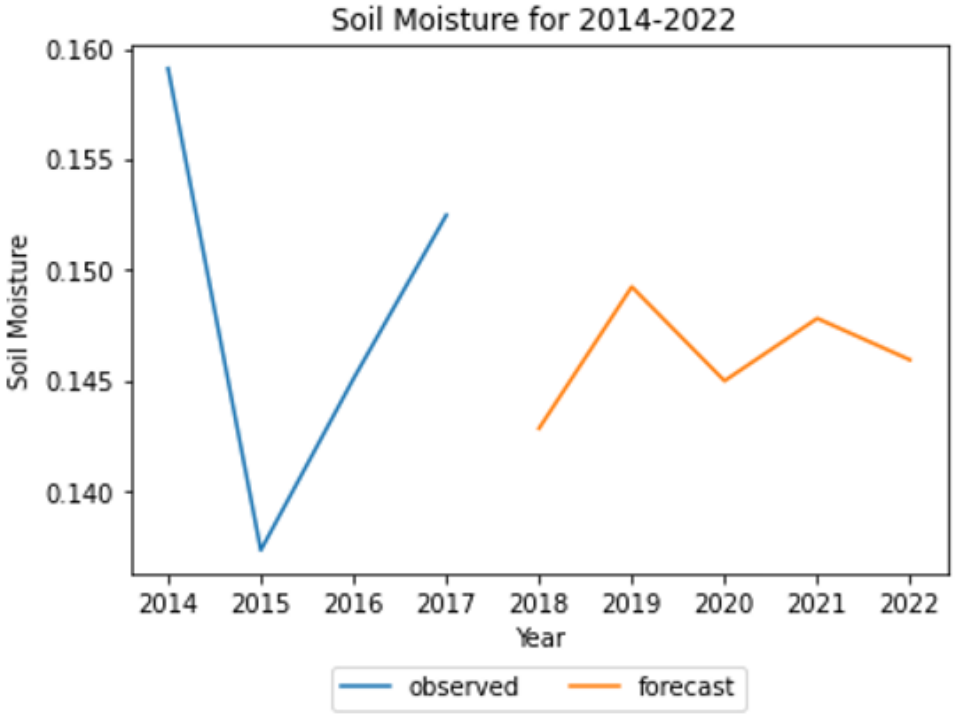


Figure 16: NDVI prediction using ARIMA



ARIMA

Figure 17: Soil moisture prediction using ARIMA

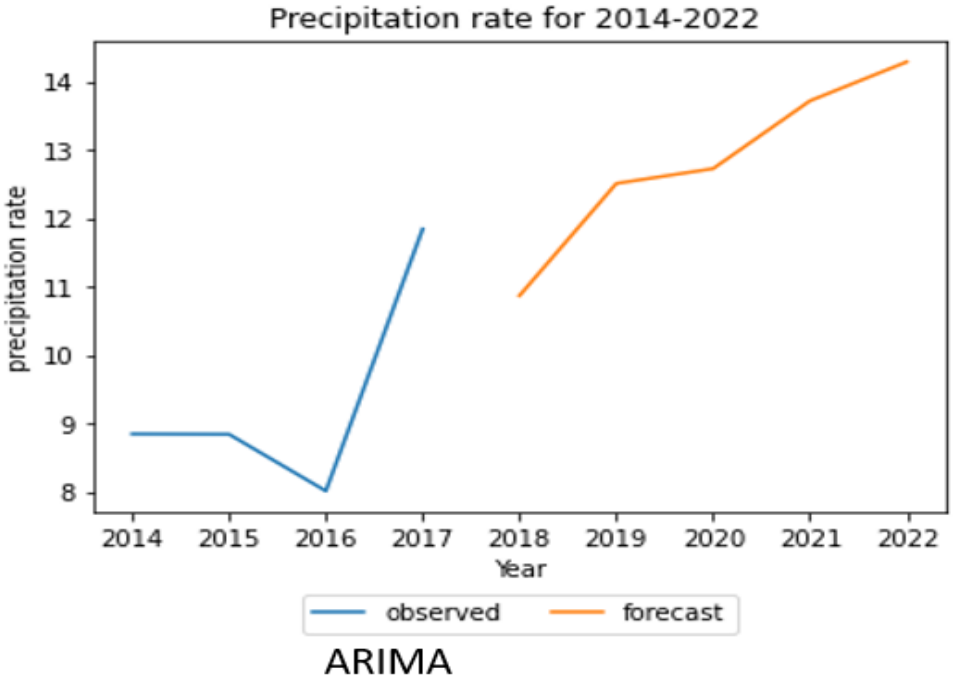


Figure 18: Precipitation rate prediction using ARIMA

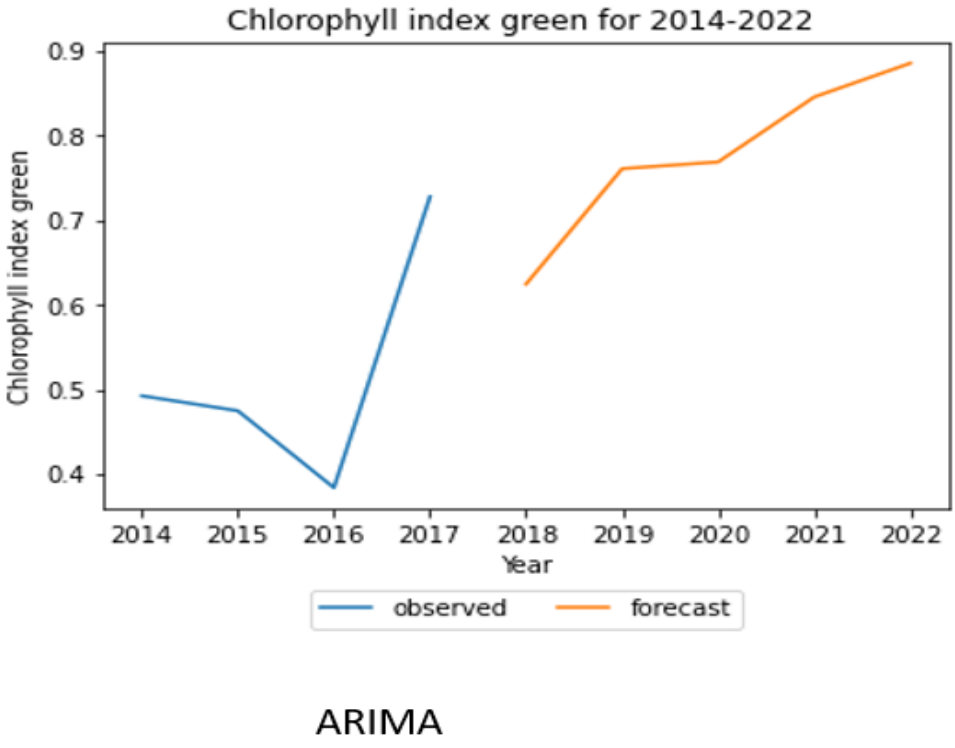


Figure 19:chlorophyll index prediction using ARIMA

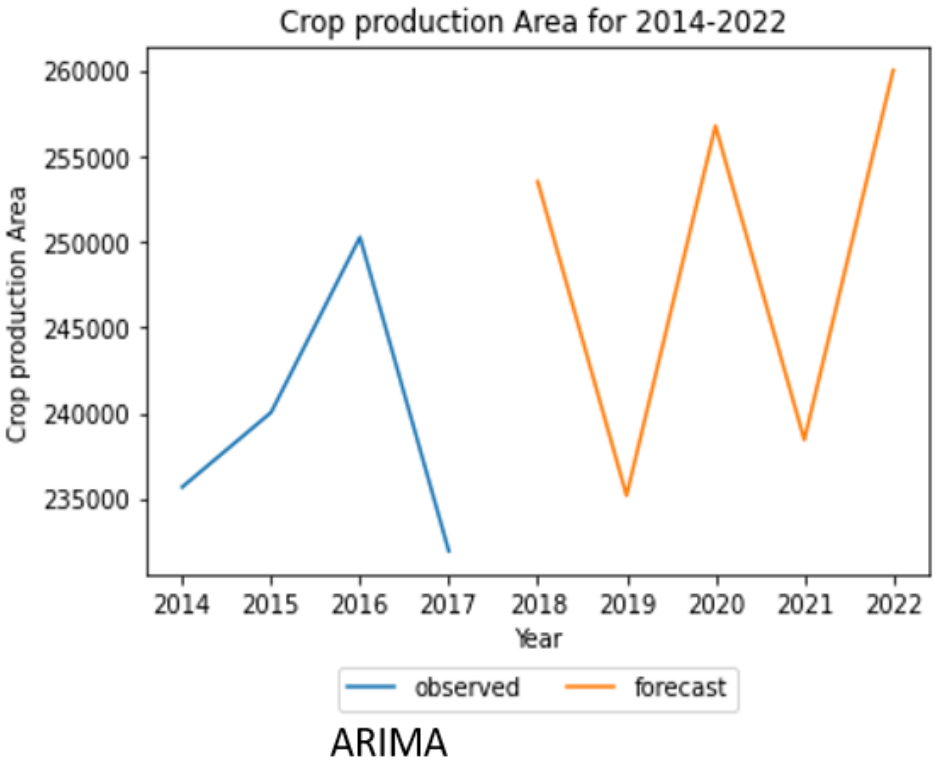


Figure 20: Crop production area prediction using ARIMA

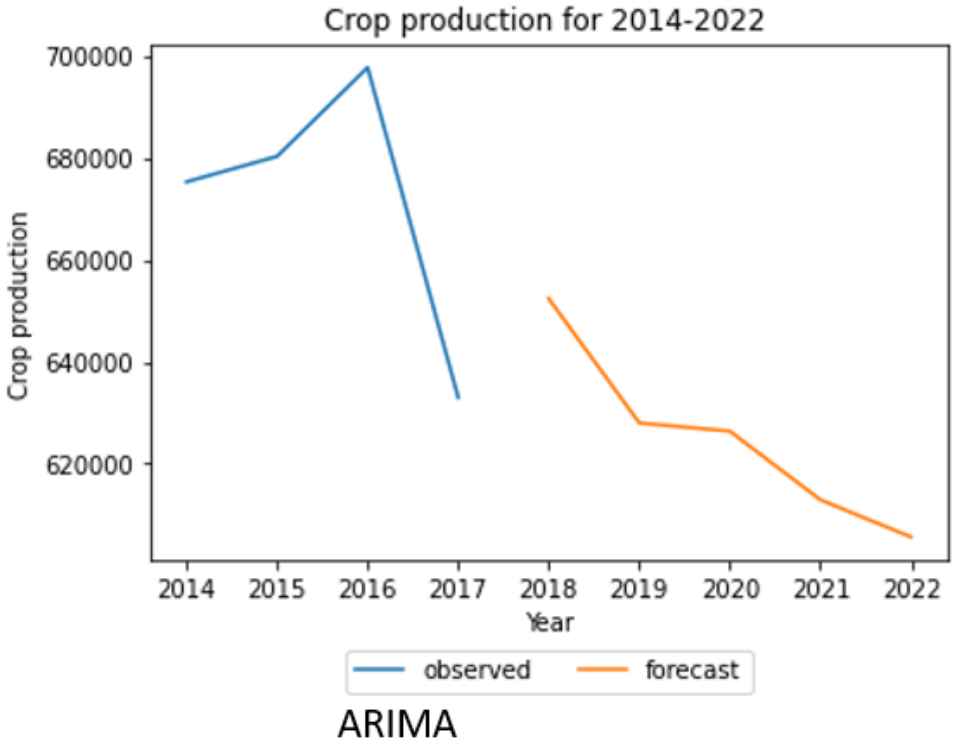


Figure 21: Crop production prediction using ARIMA

In Table 5 prediction of datasets from 2018-2022 using ARIMA model are shown.

YEAR	NDVI	NDMI	Precipitation Rate	CLG	Production Area	Crop Production
2018	0.21629542	0.14812611	10.8748796	0.624376	253535.86	652423.92
2019	0.25629563	0.14835443	12.51738848	0.76077933	235184.864	628015.88
2020	0.2356296	0.14678997	12.7342489	0.768709	256787.72	626407.78
2021	0.24630667	0.14592627	13.7270382	0.845506	238436.72	612933.41
2022	0.24079038	0.14478866	14.2975175	0.885388	260039.59	605634.86

Table 5: Prediction dataset using ARIMA

5.1.2 Crop yield datasets prediction using LSTM

Crop yield datasets prediction are also done using LSTM model. Crop yield datasets predictions are shown in figure 23,24,25,26,27 and 28.

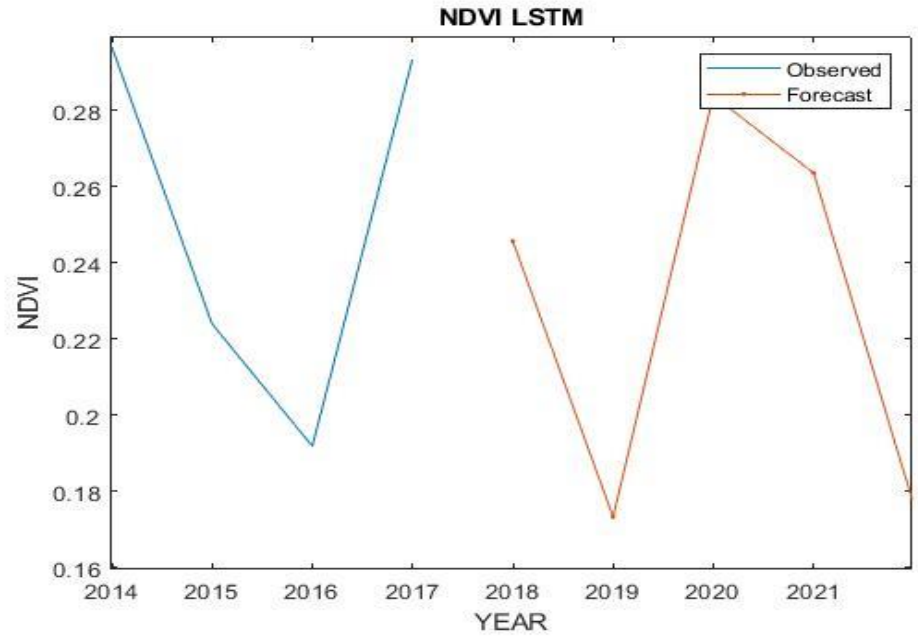


Figure 22: NDVI prediction using LSTM

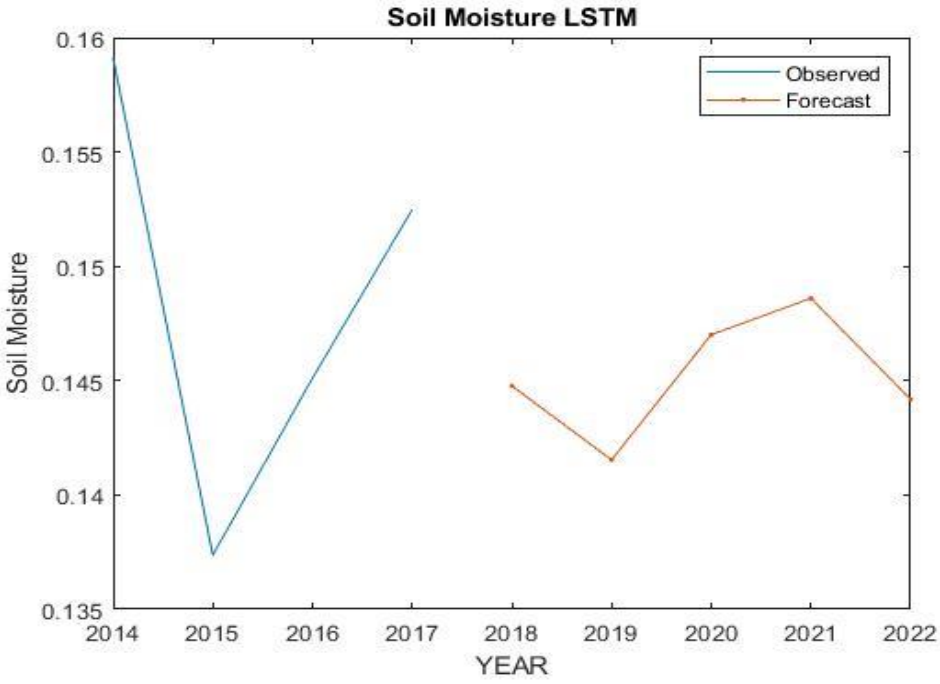


Figure 23: Soil moisture prediction using LSTM

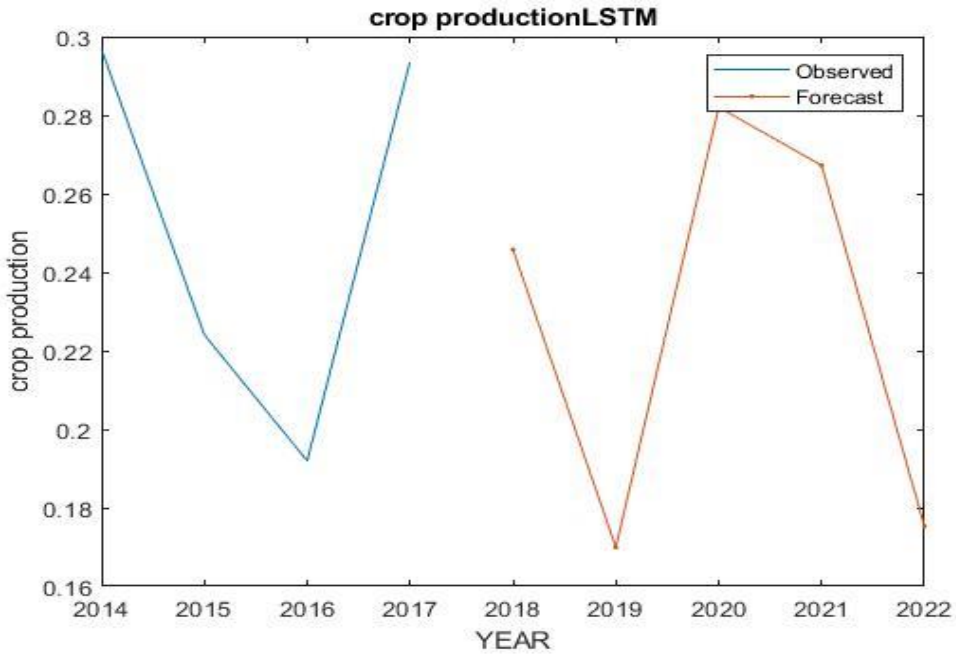


Figure 24: Crop production prediction using LSTM

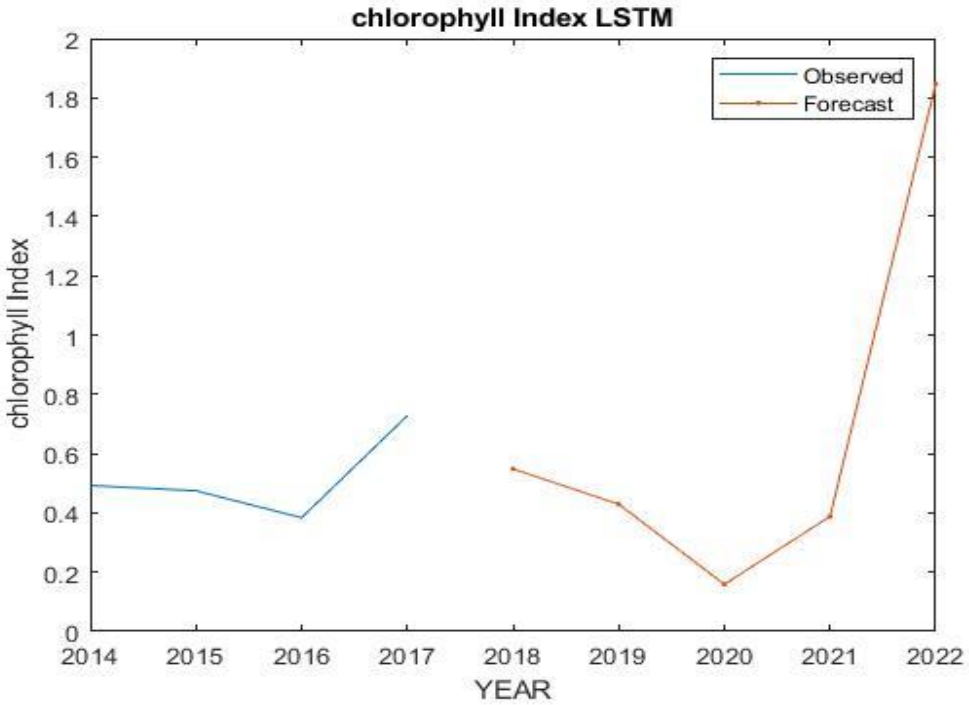


Figure 25: Chlorophyll prediction using LSTM

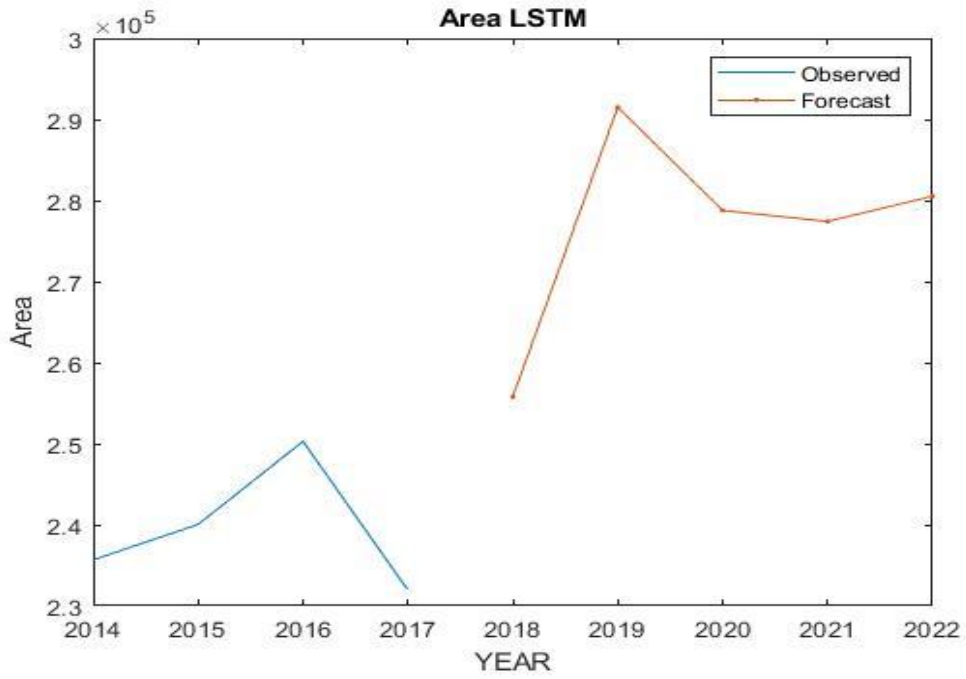


Figure 26: Area prediction using LSTM

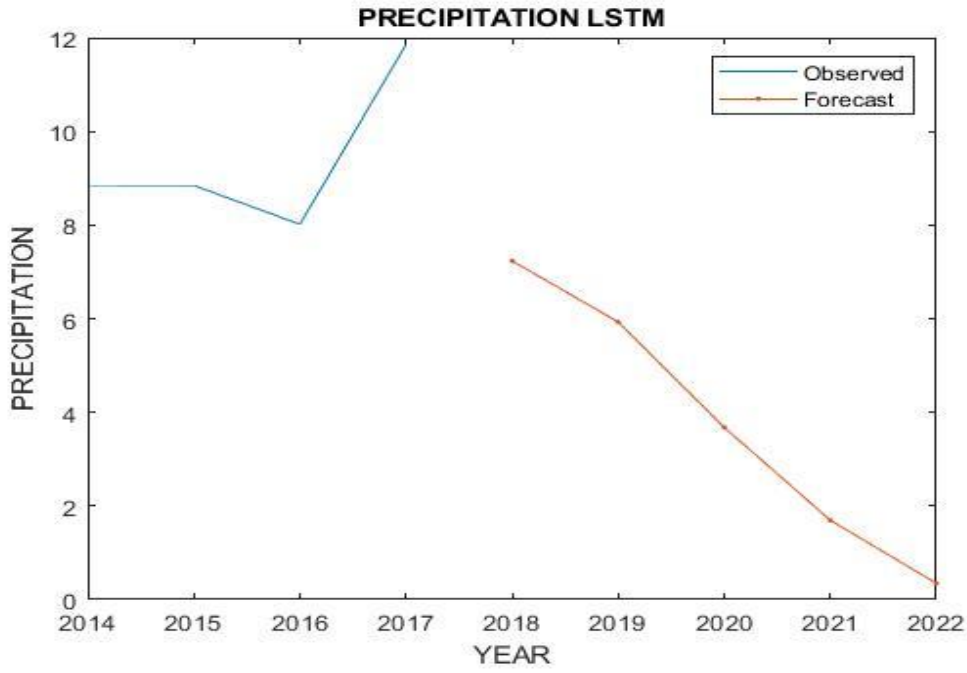


Figure 27: Precipitation prediction using LSTM

In table 6 prediction of datasets from 2018-2022 using LSTM model are shown

YEAR	NDVI	NDMI	Precipitation Rate	CLG	Production Area	Crop Production
2018	**0.24577**	**0.14477**	**7.2347**	**0.54845**	**255787**	**762660**
2019	**0.16906**	**0.1417**	**1.3443**	**0.37977**	**309407.5313**	**900355.1**
2020	**0.28431**	**0.14721**	**0.4042**	**0.17477**	**269992.9688**	**942988.9**
2021	**0.26601**	**0.14431**	**2.0553**	**0.54314**	**275126.8125**	**933217.4**
2022	**0.17207**	**0.14848**	**5.8789**	**1.3543**	**291818.84**	**930257.6**

Table 6: Prediction using LSTM

Chapter 6 Comparative study

For accuracy purpose, two types of programming platforms and models are being used in this study. A comparative study has been conducted between two programming platforms to determine the suitable one for this paper. Also, another comparative study has been concluded between two time series model for obtaining an accurate result.

6.1 Comparative Study of K-mean Segmentation using MATLAB and Python

Image segmentation classifies images into different groups. It is a very useful method for detection analysis. In this case, this algorithm classifies vegetation of Habiganj into three categories. Maximum, moderate and minimum vegetation are labeled as the three clusters and separated in both MATLAB and python. The vegetation areas are also calculated using maximum likelihood in ARCGIS software as a verification model and used for comparison in both MATLAB and Python. This study thus compares the two platforms, verifies with the area from ArcGIS, and determine which platform is suitable.

6.1.1 K-Mean clustering

In this paper, Habiganj image is segmented using K-Mean clustering. The number of clusters is defined for both the platform and they are separated and labeled as maximum, moderate and minimum vegetation. However, since the algorithm depends on the number of clusters, there is significant difference in segmentation in both the platform.

6.1.2 Computational platforms

MATLAB and Python are used to evaluate the segmentation of Habiganj images. These two platforms are highlighted because of its popularity and convenient general-purpose language.

ArcGIS uses maximum likelihood algorithm to classify the three vegetation in Habiganj. The result from ArcGIS is used as benchmark for the comparative evaluation. Both MATLAB and Python uses K-Mean algorithm to classify the vegetation types and then the corresponding areas are calculated for comparative purposes. Computational difference between MATLAB and Python is shown in figure 29.

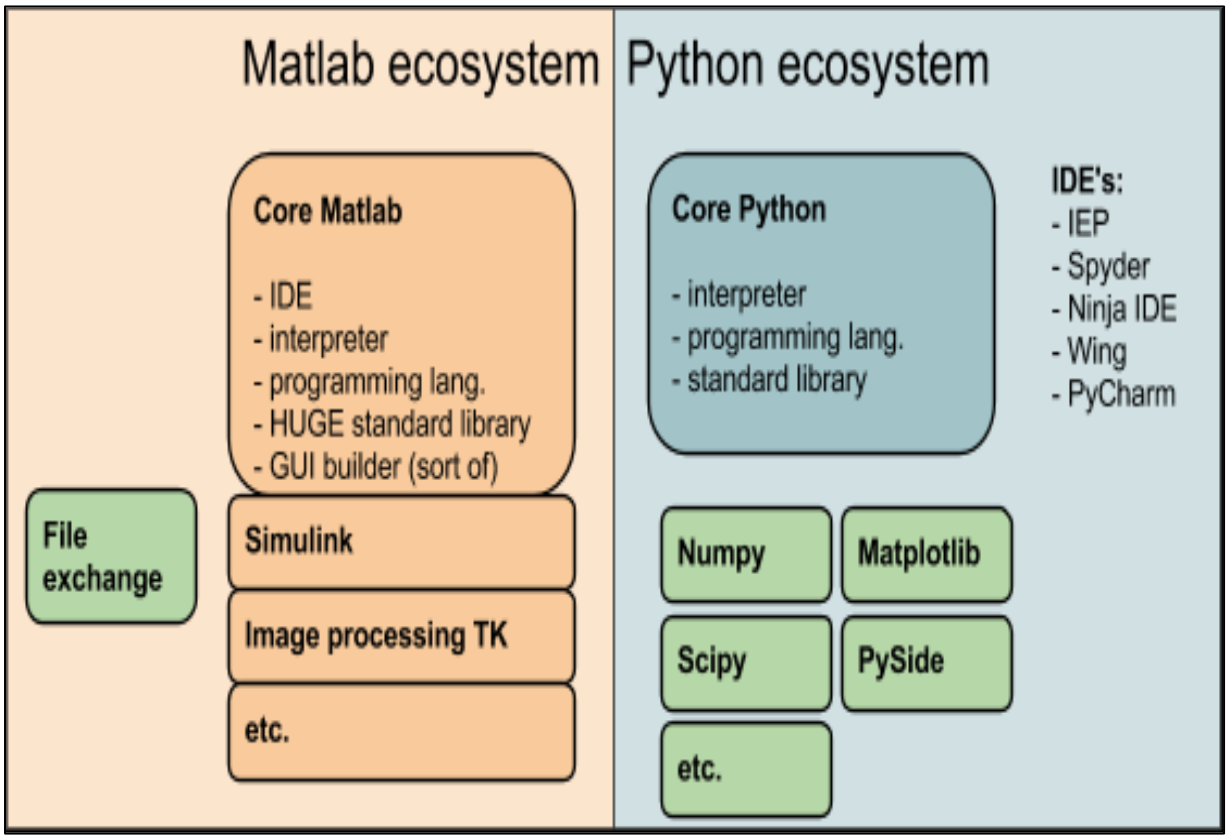


Figure 28: Python Vs MATLAB. Source [51]

6.1.3 Results

6.1.3.1 Qualitative Result

Both MATLAB and Python have used Habiganj Images and each vegetation is labeled and extracted which is shown in Figure 30. Visually it can be seen that Python shows more clear and refined areas of vegetation than MATLAB.

In high vegetation area, MATLAB has other pixel colors that might indicate overlapping of the other clusters. Python does not have such addition in the image. This indicates that MATLAB has an overlapping of the other types of vegetation resulting into a higher area percentage than Python.

Python visually seems that it has much more moderate vegetation than MATLAB. Python has more distinct colored pixels thus showing the smallest areas. It could either suggest that MATLAB does not consider all the cluster area or Python has overlapping cluster pixels.

In Lower vegetation, Python has more pixel area than MATLAB. MATLAB doesn't show a defined area and no longer maintains the boundary of Habiganj. Compared to Python, it shows a much area margin which results into higher area percentage.

The pixel area is calculated further to show the difference in the area of vegetation of Habiganj for both the platforms thus making a more rigid comparison.

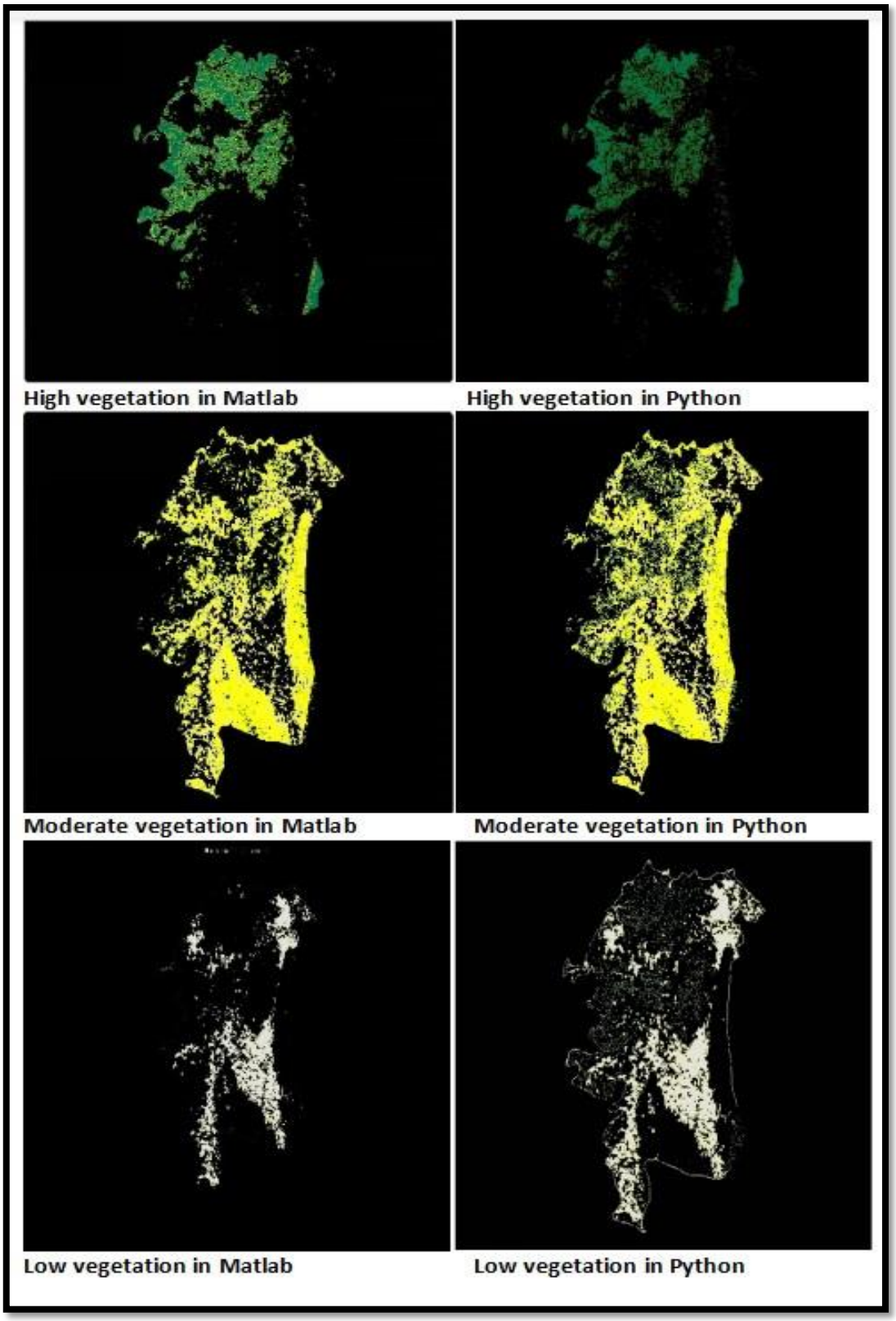


Figure 29:Difference in vegetation in MATLAB and python

6.1.3.2 Quantitative Result

With the image segmentation image result of all the three vegetation clusters, the area is calculated using both MATLAB and Python. The areas are mapped from 2015 till 2019 and plotted in graph to visualize the difference in the result. RMSE is calculated for both MATLAB and Python value to show how deviated the results from ArcGIS. It is depicted in tabular form as shown in table 7, and it can be seen that MATLAB has a considerable amount of error compare to Python.

	Root mean square error	
	MATLAB	**Python**
High Vegetation	**127.73**	**10.738**
Moderate Vegetation	**200.295**	**139.512**
Low Vegetation	**240.670**	**121.559**

Table 7: RMSE Of Python and MATLAB

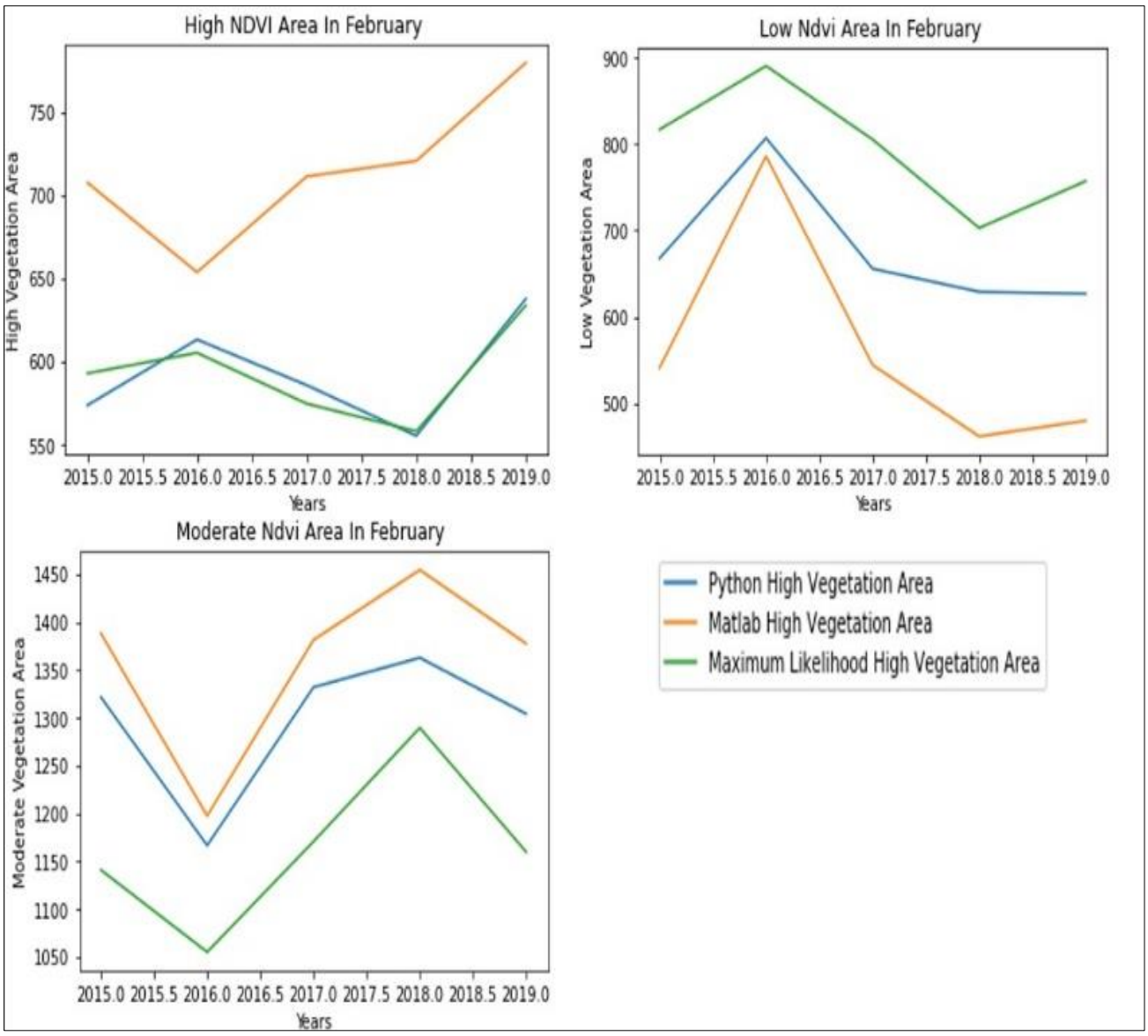


Figure 30: Quantitative Analysis

From the graphs in figure 31, it can be stated that Python generates results very close to the Maximum Likelihood algorithm in ArcGIS. So, it can be affirmed that Python is a suitable environment to carry out image segmentation of Habiganj Images compared to MATLAB.

6.2 Comparative study of crop yield parameters prediction using ARIMA and LSTM

Various datasets are utilized in crop yield prediction model. To get an accurate prediction, time series forecasting is done using ARIMA and LSTM models. These datasets are acquired by processing satellite images in ArcGIS and also retrieved from yearly database of NASA Giovanni and Bangladesh Bureau of Statistics Agriculture. In this study, comparison of the two model is verified with these obtained datasets which work as benchmark values to determine the more suitable model. Both models are popular for time series forecasting. ARIMA is a model that predicts future values based on its own past values and easier to implement and also very quick to run without needing the tuning of parameters. LSTM is a type of recurrent neural network which a set of cells which learns sequence order for prediction problems. Past value is connected to present value using these cells. This way information from past is conveyed to the present. In this model datasets are not required to be stationary but it needs a lot of data for processing and acquiring an accurate output. Thus, LSTM is more difficult to train inputs than ARIMA model. As in this study datasets from 2014-2019 have been used and LSTM is better for handling larger and complicated datasets, ARIMA is more accurate for datasets parameters. By comparing predicted test data outputs (2018-19) of ARIMA and LSTM model with benchmark values it can be determined that even though each model is working fine but ARIMA model is closer to the benchmark. Comparison between LSTM and ARIMA for different parameters are shown in figure 32,33,34,35,36 and 37.

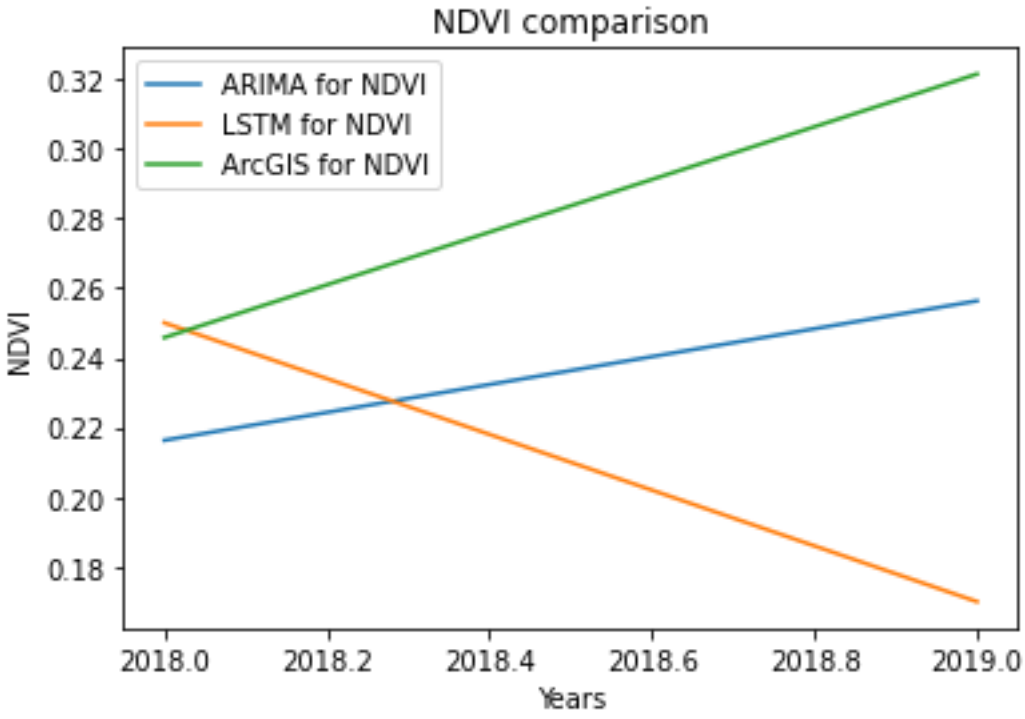


Figure 31: NDVI comparison in ARIMA, LSTM and ArcGIS

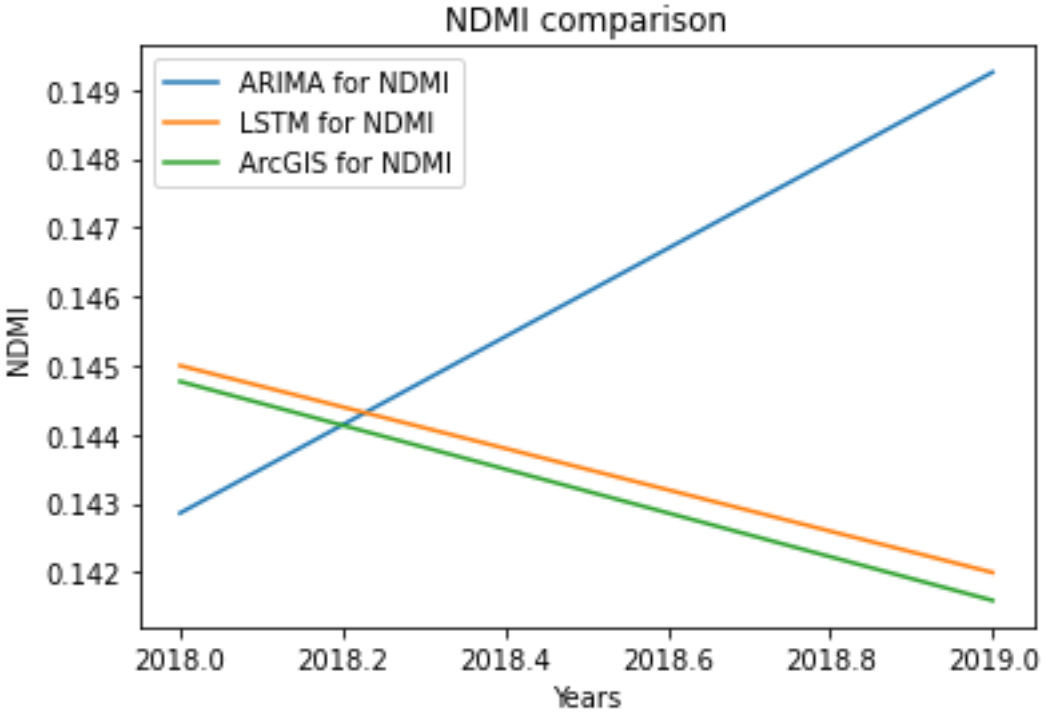


Figure 32: NDMI comparison in ARIMA, LSTM and ArcGIS

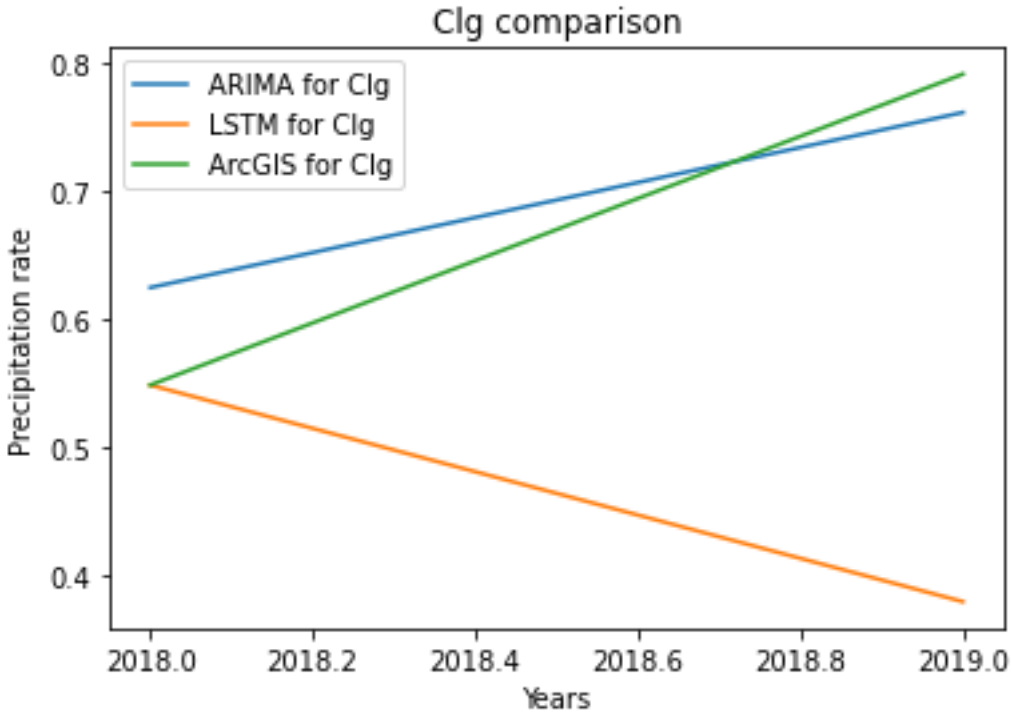


Figure 33:CLG comparison in ARIMA, LSTM and ArcGIS

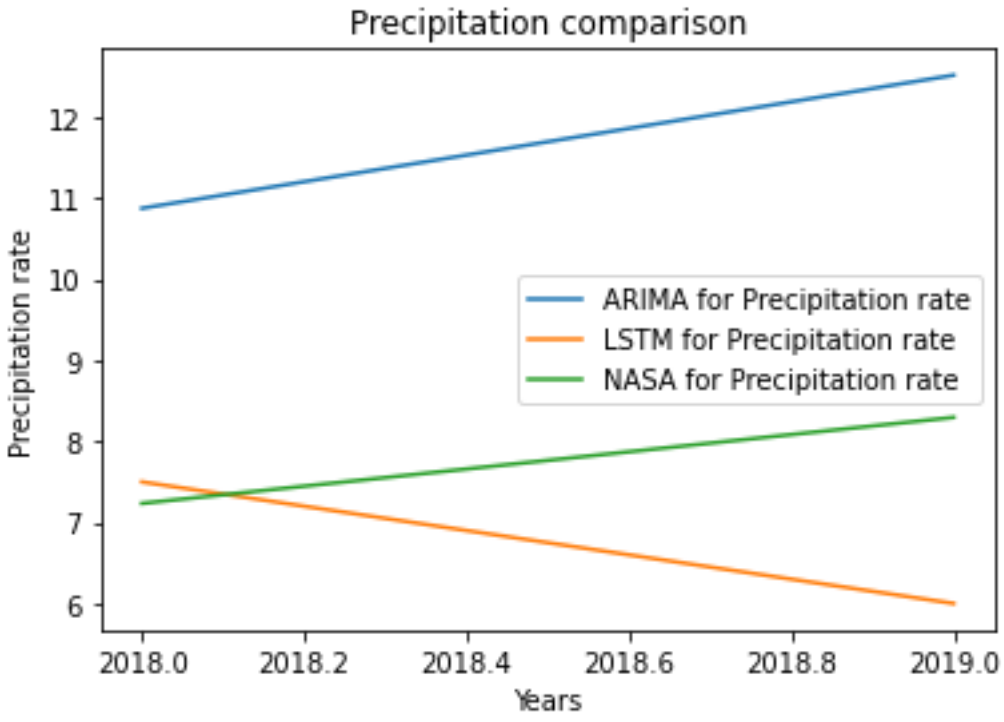


Figure 34:Precipitation comparison in ARIMA, LSTM and NASA

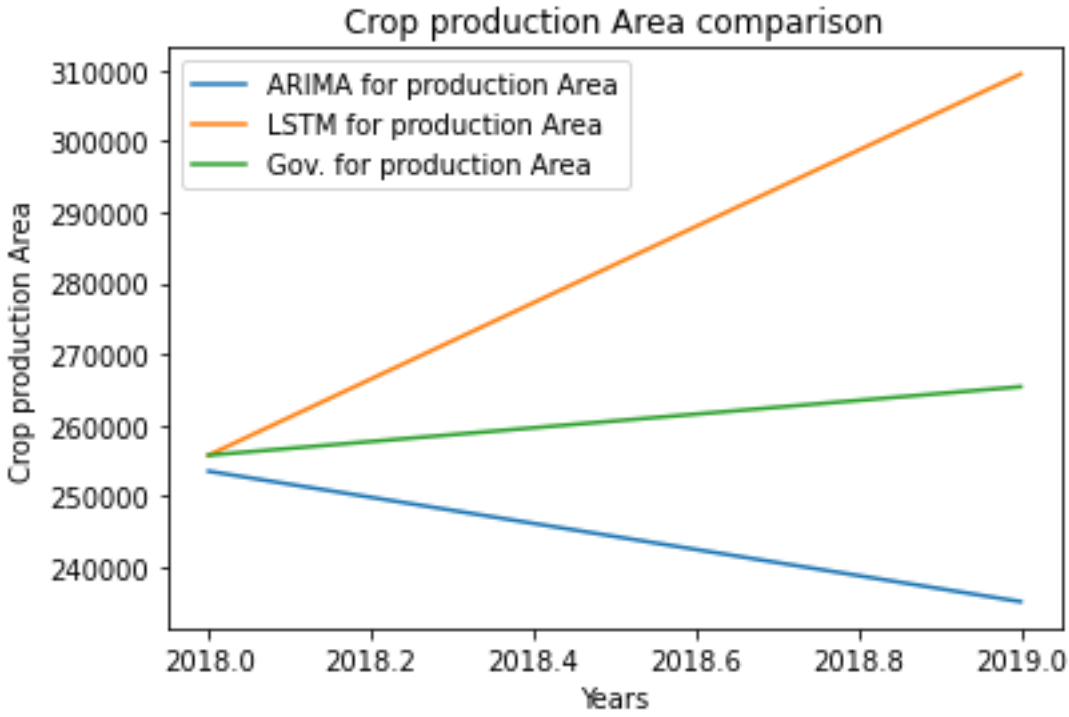


Figure 35: Crop production Area comparison in ARIMA, LSTM and Gov.

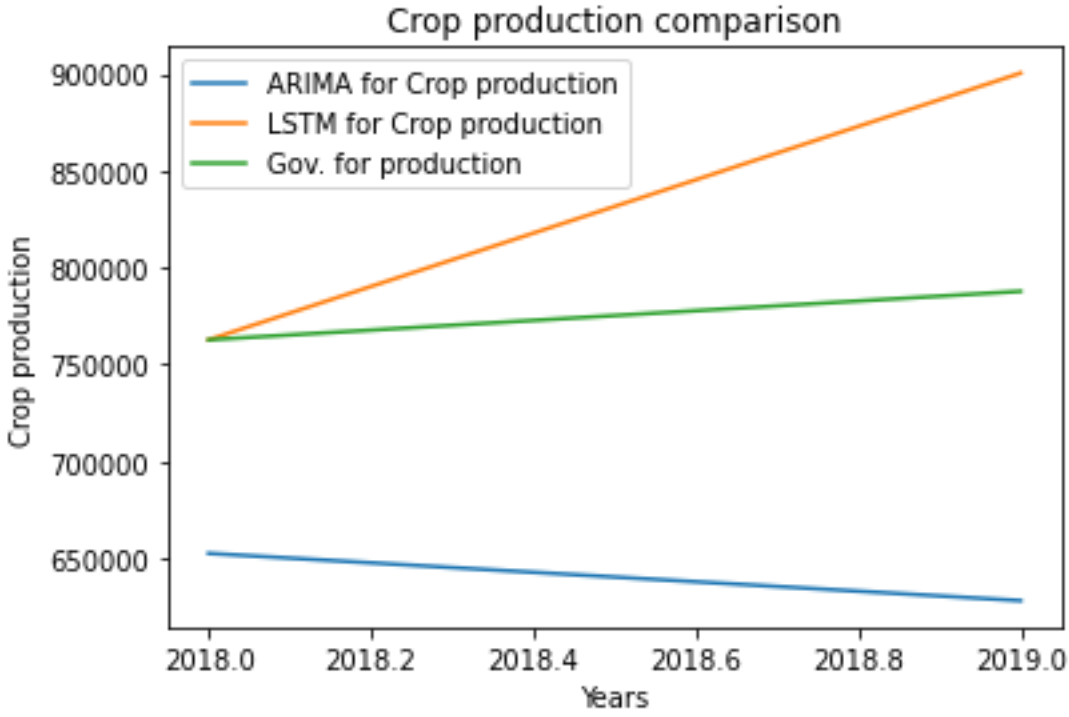


Figure 36: Crop production comparison in ARIMA, LSTM and Gov.

Furthermore, in both models RMSE is measured for accuracy and performance evaluation.

Parameter	ARIMA	LSTM
NDVI	0.0505166	0.07765
Soil Moisture	0.0055797	0.0027428
Precipitation rate	3.9415	9.9684
Chlorophyll index green	0.0577	0.43656
Production Area	21440.056	53160.1444
Crop production	137203.738	292609.54

Table 8:Accuracy and performance evaluation of ARIMA and LSTM

As it is shown in the table 8, ARIMA RMSE value is less than LSTM for all the parameters except for soil moisture (NDMI). Therefore, it is concluded that in our study ARIMA model works better than LSTM.

Chapter 7 Conclusion

7.1 Summary

In our paper, we have presented agricultural mapping and monitoring of Habiganj area along with yield prediction using satellite images. Our main goal was to develop a smart agriculture management system based on satellite imagery that would make farming efficient, profitable and manageable. Crop mapping and monitoring has been done to identify numerous zones of agricultural sector which hinders its overall efficiency and productivity. Using Landsat 8 multispectral images, we have determined land characteristics parameters like NDVI, nitrogen contents, soil moisture etc. factors that contribute to crop yield and productivity. Crop yield prediction allows to detect problems that reduces crop production. Along with remote sensing, satellite data we have used weather data such as precipitation rate and crop production area datasets for crop yield prediction. Using machine learning algorithms, like linear regression and random forest regression, future predictions on crop yield is calculated for effective decisions and as crop yield prediction depends on its datasets, predictions of these datasets using time series forecasting models ARIMA and LSTM has also been implemented. Various models have been used for crop monitoring and crop yield prediction and is compared to determine the most suited model. We have also used multiple software for crop mapping and found out that python gives more accurate result. Thus, this paper proposes a solution to analyze and monitor crop cultivation as well as predict crop yield so that agronomists and farmers can take the necessary step according to the situation beforehand. Especially considering the present COVID19 pandemic time when the financial condition is getting hugely impacted our proposed will help to boost the agricultural sector of the country as farmers will be aware of the land conditions and decisions can be made beforehand that will time efficient and economical at the same time.

7.2 Future work

In future, we will like to do more research to improve the accuracy of the smart agricultural system proposed. During our research, as mentioned earlier we used Landsat 8 satellite data. In Landsat 8 Habiganj images were available from 2014 to 2020.We could not find more than 6 years data. In future if we can get more satellite data, we can improve the accuracy of our result. Moreover, we will like to improve the accuracy and performance of the smart agricultural system using various other improved algorithm and models. In crop monitoring, NDVI is used as a health indicator and vegetation distribution in Habiganj. However, it has its limitation, so for a better evaluation and analysis, satellite that consists of red edge band. The red edge light sensor penetrates further into vegetation canopy and with its band math, provides a much broader estimation like plant disease detection and nutrition deficiencies [52].

Also, for image segmentation, besides K-Mean and Mask R-CNN other algorithms can be used be used to give a varied result that would result in a better evaluation. Ann [53] or fuzzy C [54] are one of them that can used for image segmentation which would give satisfactory results.

References

[1] A. Bangladesh, "An Overview of Agriculture in Bangladesh", *DATABD.CO*, 2020. [Online]. [Accessed: 18- July- 2020].

[2] Pandey, S, " Patterns of adoption of improved rice varieties and farm-level impacts in stress-prone rainfed areas in South Asia", International Rice Research Institute,2013. [Accessed: 18- July- 2020].

[3] Adventist Development and Relief Agency," The Plight of Farmers in Bangladesh" February 28,2019. [Accessed: 18- July- 2020].

[4] D.L. Hollinger, " Crop condition and yield prediction at the field scale with geospatial and artificial neural network applications ", Ph.D. thesis, Kent State University, Ohio, USA ,2011

[5] A. Gonzalez-Sanchez, J. Frausto-Solis and W. Ojeda-Bustamante, "Predictive ability of machine learning methods for massive crop yield prediction", Spanish Journal of Agricultural Research, vol. 12, no. 2, p. 313, 2014.

[6] H. Cheng, H. Peng and S. Liu, "An improved K-means clustering algorithm in agricultural image segmentation",2020.

[7] B. Gudmundsdottir, "Detection of potential arable land with remote sensing and GIS: a case study for Kjósarhreppur", Lup.lub.lu.se, 2020. [Accessed: 23- Jun- 2020].

[8] "What is precision farming and how to get started with it", Blog.onesoil.ai[online]. [Accessed: 22- Jun- 2020].

[9] "Precisionagriculture", En.wikipedia.org [Accessed: 23- Aug- 2020].

[10] Pauline Chivenge, Sheetal Sharma "Precision Agriculture In Food Production: Nutrient Management"Presented At International Workshop On Icts For Precision Agriculture, Mardi Headquarters, Selangor, Malaysia 6 - 8 August 2019

[11] ”PRECISION AGRICULTURE: FROM CONCEPT TO PRACTICE”,eos.com,FEBRUARY 22, 2019. [Accessed: 24- Sep- 2020].

[12] Mahfuza Afroj, Mohammad Mizanul Haque Kajol ,Mahfuzar Rahman ,”Precision agriculture in the world and its prospect in Bangladesh,”Res.Agric.Livest.Fish.,vol.3,No 1,pp.01-14, 2016

[13] Elizabeth Howell ,” What is a Satellite?”, October 27, 2017. [Accessed: 24- Sep- 2020].

[14] Dan Stillman,” What Is a Satellite?”, www.nasa.gov, Feb. 12, 2014. [Accessed: 24- Sep- 2020].

[15] Colin R. Leslie, Larisa O. Serbina, and Holly M. Miller,” Landsat and Agriculture—Case Studies on the Uses and Benefits of Landsat Imagery in Agricultural Monitoring and Production”, presented at U.S. Geological Survey, Reston, Virginia: 2017.

[16] “Landsat 8 Overview « Landsat Science", *Landsat.gsfc.nasa.gov*, 2020. [Online]. [Accessed: 24- Sep- 2020].

[17] “[10]"Landsat 8", *Usgs.gov*, 2020. [Online]. [Accessed: 24- Sep- 2020].

[18] M. A. Ridwan, N. A. M. Radzi, W. S. H. M. W. Ahmad, I. S. Mustafa, N. M. Din, Y. E. Jalil, A. M. Isa, N. S. Othman, W. M. D. W. Zaki,” Applications of Landsat-8 Data: a Survey”, January 2018

[19] " What are the band designations for the Landsat satellites?", *Usgs.gov*, 2020. [Online]. [Accessed: 24- Sep- 2020].

[20] Bangladesh Bureau of Statistics (BBS) ,“District Statistics 2011 Habiganj” December, 2013.Available:http://203.112.218.65:8008/WebTestApplication/userfiles/Image/District%20Statistics/Habiganj.pdf

[21] Mohammad Nur Uddin,” Hybrid rice production sets new record in Habiganj” Dhaka tribune.com,17thApril2019.

[22] Pandey, S, Patterns of adoption of improved rice varieties and farm-level impacts in stress-prone rainfed areas in South Asia: International Rice Research Institute,2013

[23] "What is MATLAB?", *Mathworks.com*,2020. [Accessed: 20- Sep- 2020].

[24] "What is Python? Executive Summary", *Python.org*, 2020. [Accessed: 20- Sep- 2020].

[25] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN", *IEEE International Conference on Computer Vision*, 2017.

[26] N. Dhanachandra, K. Manglem and Y. Chanu, "Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm", *Procedia Computer Science*, vol. 54, pp. 764-771, 2015.

[27] M. Sonawane and C. Dhawale, "A Brief Survey on Image Segmentation Methods", *International Journal of Computer Applications(0*, 2020).

[28] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN", *IEEE International Conference on Computer Vision*, 2017.

[29] Konstantinos G. Liakos, , Patrizia Busato , Dimitrios Moshou, Simon Pearson and Dionysis Bochtis," Machine Learning in Agriculture: A Review", 14August2018

[30] "Long short term memory", *Mathworks.com*,2020. [Accessed: 24- Sep- 2020].

[31] "EarthExplorer", *Earthexplorer.usgs.gov*, 2020. [Accessed: 21- Sep- 2020].

[32] ArcMap | Documentation", *Desktop.arcgis.com*, 2020. [Online]. [Accessed: 21- Sep- 2020].

[33] Biljana Bojovic , Aca Markovic," CORRELATION BETWEEN NITROGEN AND CHLOROPHYLL CONTENT IN WHEAT (Triticum aestivum L.)", March 19, 2009

[34] "Landsat Normalized Difference Vegetation Index", *Usgs.gov*, 2020. [Accessed: 22- Sep- 2020].

[35] C. Qiu *et al.*, “Derivative Parameters of Hyperspectral NDVI and Its Application in the Inversion of Rapeseed Leaf Area Index,” *Applied Sciences*, vol. 8, no. 8, p. 1300, Aug. 2018.

[36] K.-A. Nguyen, Y.-A. Liou, H.-P. Tran, P.-P. Hoang, and T.-H. Nguyen, “Soil salinity assessment by using near-infrared channel and Vegetation Soil Salinity Index derived from Landsat 8 OLI data: a case study in the Tra Vinh Province, Mekong Delta, Vietnam,” *Prog Earth Planet Sci*, vol. 7, no. 1, p. 1, Dec. 2020

[37] "Normalized Difference Moisture Index", *Usgs.gov*, 2020. [Online]. [Accessed: 22- Sep- 2020].

[38] B. J. Buenaobra, M. V. Manhuyod, and R. E. Otadoy, “VISUALIZATION, PROFILING AND CALIBRATION OF SELECTED VEGETATION INDICES FROM SUBSETTED AND PANCHROMATIC BAND SHARPENED LANDSAT IMAGERY USING FREE AND OPEN SOURCE SOFTWARE GIS/RS TOOLS,” p. 6.

[39] Bannari, A., Morin, D. and Bonn, F. (1995) 'A review of vegetation indices', Remote Sensing Reviews, Vol. 13, pp. 95-120,1995

[40] T. Carlson and D. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index", *Remote Sensing of Environment*, vol. 62, no. 3, pp. 241-252, 1997.

[41] *OpenCV*.org. [Online]. https://opencv.org/about/. [Accessed: 18- Sep- 2020]

[42] X. Zheng, Q. Lei, R. Yao, Y. Gong and Q. Yin, "Image segmentation based on adaptive K-means algorithm", *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, 2018.

[43] Bbs.gov.bd, 2020. [Accessed: 13- Sep- 2020].

[44] "Average Weather in Habiganj, Bangladesh, Year-Round - Weather Spark", *Weatherspark.com*, 2020. [Online]. [Accessed: 21- Sep- 2020].

[45] Dam collapses, villages flooded in Habiganj", *Dhaka Tribune*, 2020. [Accessed: 21- Sep- 2020].

[46] R. Machado and R. Serralheiro, "Soil Salinity: Effect on Vegetable Crop Growth. Management Practices to Prevent and Mitigate Soil Salinization", *Horticulturae*, vol. 3, no. 2, p. 30, 2017.

[47] A. Gonzalez-Sanchez, J. Frausto-Solis, and W. Ojeda-Bustamante, "Predictive ability of machine learning methods for massive crop yield prediction," *Span J Agric Res*, vol. 12, no. 2, p. 313, Apr. 2014

[48] A. Burkov, The Hundred-Page Machine Learning Book.

[49] M. Dastorani, Mirzavand, Dastorani and Sadatinejad, "Comparative study among different time series models applied to monthly rainfall forecasting in semi-arid climate condition", *Ideas.repec.org*, 2020. [Accessed: 24- Sep- 2020].

[50] N. IQBAL, K. BAKHSH, A. MAQBOOL and A. AHMAD, "Use of the ARIMA Model for Forecasting Wheat Area and Production in Pakistan", *https://www.researchgate.net*, 2020.

[51] "Pyzo - python_vs_matlab", *Pyzo.org*, 2020. [Accessed: 20- Sep- 2020].

[52] B. Boiarskii, "Comparison of NDVI and NDRE Indices to Detect Differences in Vegetation and Chlorophyll Content", *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES*, vol. 1, no. 4, 2019.

[53] M. Jayanthi and D. Shashikumar, "Survey on Agriculture Image Segmentation Techniques", *Asian Journal of Applied Science and Technology (AJAST)*, vol. 1, no. 8, pp. 143-146, 2017. [Accessed 19 September 2020].

[54] S. Madhukumar and N. Santhiyakumari, "Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain", *The Egyptian Journal of Radiology and Nuclear Medicine*, vol. 46, no. 2, pp. 475-479, 2015.

Appendix A

Code for Crop yield prediction in python

```
1. x=data.iloc[:, :-1]
2. y=data.iloc[:,7]


#splittingDatasetIntoTrainingSetandTestingData
3. from sklearn.model_selection import train_test_split
4. x_train, x_test, y_train, y_test= train_test_split(x,y, test_size = 0.2,random_state= 0)


 #linear_regression
 5. from sklearn.linear_model import LinearRegression
 6. regressor= LinearRegression()
 7. regressor.fit(x_train, y_train)

  #predicting the testset results

 8. y_pred=regressor.predict(x_test)
 9. from sklearn.metrics import r2_score
 10. score=r2_score(y_test,y_pred)


  # absolute errors
  11. import sklearn.metrics as metrics
  12. from sklearn.metrics import mean_squared_error
  13. print('MAE:', metrics.mean_absolute_error(y_test, y_pred))
  # Calculate mean absolute percentage error (MAPE)
  14. mape = 100 * (metrics.mean_absolute_error(y_test, y_pred) / y_test)

  # accuracy
  15. accuracy = 100 - np.mean(mape)
  16. print('Accuracy:', round(accuracy, 2), '%.')

  #mean squared error
  17. print('MSE:', metrics.mean_squared_error(y_test, y_pred))
  # Calculate mean squared percentage error (MSPE)
  18. mspe = 100 * (metrics.mean_squared_error(y_test, y_pred) / y_test)
  # Calculate and display accuracy
  19. accuracy = 100 - np.mean(mspe)
  20. print('Accuracy:', round(accuracy, 2), '%.')

  #Root mean squared error for linear regression
  21. print('RMSE:',np.sqrt(mean_squared_error(y_test,y_pred)))
  22. rmspe = 100 * (metrics.mean_squared_error(y_test, y_pred) / y_test)
```

```
23. accuracy=100-(np.sqrt(mean_squared_error(y_test,y_pred))/y_test)
# Calculate and display accuracy
24. accuracy = 100 - np.mean(rmspe)
25. print('Accuracy:', round(accuracy, 2), '%.')


 #import pickle
 26. from sklearn.externals import joblib
 27. joblib.dump(regressor, "multiple_regression_model.pkl")

 #predicting the single observation results using linear regression for 2022

 28. from sklearn.externals import joblib

 29. YEAR=2022

 30. Soil_moisture=0.14478866
 31. NDVI=0.24079038
 32. Precipitation_rate=14.2975175

 33. Area=260039.592
 34. Production= 605634.867
 35. Chlorophyll_index= 0.88538812
 36. prediction_data=[YEAR,Soil_moisture,NDVI,Precipitation_rate,Area,
 Production, Chlorophyll_index]
 37. prediction_data_array=np.array(prediction_data)
 38. prediction_data_array=prediction_data_array.reshape(1,-1)
 39. model=open("multiple_regression_model.pkl","rb")
 40. prediction_model=joblib.load(model)
 41. print(prediction_data_array.size)
 42. model_prediction=pred_model.predict(prediction_data_array)
 43. round(float(model_predictioction),2)


#RandomForestRegression

#Scaling
44. from sklearn.preprocessing import StandardScaler

45. sc = StandardScaler()
46. x_train = sc.fit_transform(x_train)
47. x_test = sc.transform(x_test)
48. from sklearn.ensemble import RandomForestRegressor

49. regressor = RandomForestRegressor(n_estimators=4, random_state=0)
50. regressor.fit(x_train, y_train)
51. y_pred = regressor.predict(x_test)
```

For calculating errors and single observation result for year 2022 in random forest regression, same code is used as linear regression.

Code for dataset prediction using ARIMA model

```
#ARIMA - Autoregressive(p)Integrated(d)Moving Average(q)
#for NDVI
1. from statsmodels.tsa.arima_model import ARIMA
2. from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

3. plot_acf(ndvi)

4. plot_pacf(ndvi)

5. veg_train= ndvi[0:4]

6. veg_test=ndvi[4:6]
7. veg_model= ARIMA(veg_train, order=(1,1,0))

8. veg_model_fit=veg_model.fit()

9. veg_model_fit.aic

10.veg_forcast=veg_model_fit.forecast(steps=5)[0]

11. veg_forcast
```

Using this same code with different datasets' value, all datasets is predicted.

Code for k mean clustering in MATLAB

```
I=imread('ndvi_f15.jpg');
imshow(I)

lab_he = rgb2lab(I);

ab = lab_he(:,:,2:3);
ab = im2single(ab);
nColors = 3;

pixel_labels = imsegkmeans(ab,nColors,'NumAttempts',3);
figure
imshow(pixel_labels,[])
title('Image Labeled by Cluster Index')


mask1 = pixel_labels==1;
cluster1 = I .* uint8(mask1);
figure
imshow(cluster1)

title('Objects in Cluster 1');

mask2 = pixel_labels==2;
cluster2 = I .* uint8(mask2);
figure
imshow(cluster2)

title('Objects in Cluster 2');


mask3 = pixel_labels==3;
cluster3 = I .* uint8(mask3);
figure
imshow(cluster3)

title('Objects in Cluster 3');
```

Code for area calculation from k mean pixel count

For minimum and moderate

```
I=imread('j16modp.jpg');%reading image
figure
imshow(I)%displaying image
g=rgb2gray(I);%converting to grey
imshow(g)%displaying image
B = im2bw(g);%converting grey to binary
```

```
figure

imshow(B)
numberOfPixels = numel(B);%counting total pixels
numberOfTruePixels = sum(B(:));%counting total white pixels
B=numberOfPixels-numberOfTruePixels;%counting total black pixels
```

For maximum

```
I=imread('f15maxP.jpg');%reading image
figure
imshow(I)%displaying image
g=rgb2gray(I);%converting to grey

figure
imshow(g)%displaying image g

level=0.09;
thres=im2bw(g,level); %threshholding for the trapped yellow parts
figure
imshow(thres);%displaying image g
numberOfPixels = numel(thres);%counting total pixels
numberOfTruePixels = sum(thres(:));%counting total white pixels
B=numberOfPixels-numberOfTruePixels;%counting total black pixels
```

Code for dataset prediction of parametrs using LSTM

All the parameters prediction is done using this code for LSTM time series.

```
data =PARAMETER;

data = [data{:}];
figure
plot(data)
xlabel("YEAR")
xticks([1 2 3 4 5 6 7 8 9])
xticklabels({2014 2015  2016 2017 2018 2019 2020 2021 2022})
ylabel("PARAMETER")
title("PARAMETER TIMES SERIES")


num_steps_train=floor(.9*numel(data)) ;
train=data(1:num_steps_train);
test=data(num_steps_train-3:end);

sig = std(train);
```

```
std_train_data=(train-mu)/sig;


XTrain = std_train_data(1:end-1);
YTrain =std_train_data(2:end);


numFeatures = 1;
numResponses = 1;
numHiddenUnits = 200;

layers = [
    sequenceInputLayer(numFeatures)
    lstmLayer(numHiddenUnits)
    fullyConnectedLayer(numResponses)
    regressionLayer];

options = trainingOptions('adam', ...
    'MaxEpochs',250, ...
    'GradientThreshold',1, ...
    'InitialLearnRate',0.005, ...
    'LearnRateSchedule','piecewise', ...
    'LearnRateDropPeriod',125, ...
    'LearnRateDropFactor',0.2, ...
    'Verbose',0, ...
    'Plots','training-progress');


net = trainNetwork(XTrain,YTrain,layers,options);


std_train_data = (test - mu) / sig;
XTest = std_train_data(1:end-1);

net = predictAndUpdateState(net,XTrain);
[net,YPred] = predictAndUpdateState(net,YTrain(end));

numTimeStepsTest = numel(XTest);
for i = 2:numTimeStepsTest
    [net,YPred(:,i)] = predictAndUpdateState(net,YPred(:,i-1),'ExecutionEnvironment','cpu');
end
YPred = sig*YPred + mu;

YTest = test(2:end);
rmse = sqrt(mean((YPred-YTest).^2));


plot(train(1:end-1))
hold on
idx = num_steps_train:(num_steps_train+numTimeStepsTest);
```

```
plot(idx,[data(num_steps_train) YPred],'.-')
hold off


xlabel("YEAR")
xticks([1 2 3 4 5 6 7 8 9])
xticklabels({2014 2015  2016 2017 2018 2019 2020 2021 2022})
ylabel("PARAMETER")
title("PARAMETER LSTM")

legend(["Observed" "Forecast"])

figure
subplot(2,1,1)
plot(YTest)
hold on
plot(YPred,'.-')
hold off
legend(["Observed" "Forecast"])
ylabel("Area")
xticks([1 2 3 4])
xticklabels({ 2016 2017 2018 2019 })
title("Forecast")

subplot(2,1,2)
stem(YPred - YTest)
xlabel("YEAR")
xticks([1 2 3 4])
xticklabels({ 2016 2017 2018 2019 })
ylabel("Error")
title("RMSE = " + rmse)

net = resetState(net);
net = predictAndUpdateState(net,XTrain);

YPred = [];
numTimeStepsTest = numel(XTest);
for i = 1:numTimeStepsTest
    [net,YPred(:,i)] = predictAndUpdateState(net,XTest(:,i),'ExecutionEnvironment','cpu');
end
YPred = sig*YPred + mu;

rmse = sqrt(mean((YPred-YTest).^2));


figure
subplot(2,1,1)
plot(YTest)
hold on
plot(YPred,'.-')
```

```
hold off
legend(["Observed" "Predicted"])
ylabel("Area")
xticks([1 2 3 4])
xticklabels({ 2016 2017 2018 2019 })
title("Forecast with Updates")

subplot(2,1,2)
stem(YPred - YTest)
xlabel("YEAR")
xticks([1 2 3 4])
xticklabels({ 2016 2017 2018 2019 })
ylabel("Error")
title("RMSE = " + rmse)
```

Variable parameter is replaced for each case

Code for K-Mean Algorithm in Python for Healthy Vegetation

```
import cv2

import numpy as np

import matplotlib.pyplot as plt

#Import the images from directory

im=cv2.imread(r"C:\Users\Personal\Desktop\NDVI_V2-20200713T172249Z-
001\NDVI_V2\ndvi\ndvi_nov18.jpg")

pixel_value = im.reshape((-1, 3))

#Converting the image to float

im2 = np.float32(pixel_value)

criteria = (cv2.TERM_CRITERIA_EPS + cv2.TERM_CRITERIA_MAX_ITER, 100, 0.2)

#Choose appropriate cluster value k

k = 6

ret, labels, (centers) = cv2.kmeans(im2, k, None, criteria, 10,
cv2.KMEANS_RANDOM_CENTERS)

#Convert back to 8 bit

centers = np.uint8(centers)

#Label array is flattened

labels = labels.flatten()

segmented_image = centers[labels]
```

```
segmented_image = segmented_image.reshape(im.shape)

#Display

plt.imshow(segmented_image)

plt.show()


#Masking all  vegetation except healthy vegetation

masked_image = np.copy(im)

# Convert to the shape of a vector of pixel values

masked_image = masked_image.reshape((-1, 3))


#Choose any cluster label number to mask it to a black background

cluster =

masked_image[labels == cluster] = [0, 0, 0]

# Convert back to original size

masked_image = masked_image.reshape(im.shape)

# Display and save the image

plt.imshow(masked_image)

plt.show()

cv2.imwrite( "Nov18.jpg",masked_image )
```

Code for area calculation for images in MATLAB

```
import cv2

import numpy as np

#Pixel Area of white pixels from a gray image

image = cv2.imread(r'C:\Users\Personal\Desktop\Comparative study\Python\Feb19_high veg.jpg')

gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

thresh = cv2.threshold(gray,0,255,cv2.THRESH_OTSU + cv2.THRESH_BINARY)[1]

cnts = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)

cnts = cnts[0] if len(cnts) == 2 else cnts[1]

high = 0


for c in cnts:

    x,y,w,h = cv2.boundingRect(c)

    mask = np.zeros(image.shape, dtype=np.uint8)

    cv2.fillPoly(mask, [c], [255,255,255])

    mask = cv2.cvtColor(mask, cv2.COLOR_BGR2GRAY)

    pixels = cv2.countNonZero(mask)

    high += pixels
```

```
    cv2.putText(image, '{}'.format(pixels), (x,y - 15), cv2.FONT_HERSHEY_SIMPLEX, 0.8,
(255,255,255), 2)

print(high)cv2.waitKey(0)
```

#Pixel area in sqkm

```
Habiganj_area=2637

Habiganj_Pixel_Area=high+mid+low

print("Habiganj pixel count",Habiganj_Pixel_Area)

pixel_per_km=2637/Habiganj_Pixel_Area

Low_Veg=pixel_per_km*low

Mid_Veg=pixel_per_km*mid

High_Veg=pixel_per_km*high

print(f"The High Vegetation Area is {High_Veg:.2f} sqkm")

print(f"The Moderate Vegetation Area is {Mid_Veg:.2f} sqkm")

print(f"The Low Vegetation Area is {Low_Veg:.2f} sqkm")
```